



DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
1400 WILSON BOULEVARD
ARLINGTON VIRGINIA 22209

IPTO/Carls
L
#3225
Reproduced at the National Archives

ARPA Order No. 3759
Program Code No. 9P10
Program Element Code: 62706E
Industrial Priority Rating: D0

March 13, 1979

Director for Budget & Finance
Washington Headquarters Services
Washington, D.C. 20301

1. This Amendment provides \$75,000 for issuance to the Defense Supply Services - Washington (DSS-W) to augment the funding of Contract No. MDA903-76-C-0260 with Systems Development Corporation (SDC) in support of the DARPA/IPTO KVM370 Security program. The funds were allocated to DARPA for this jointly funded effort under CIA MIPR 900907, dated 27 February 1979.
2. This research is a continuation of work previously performed under ARPA Order No. 3225 via the above cited contract and is for approximately three (3) months effort on KVM370 Integration Testing.
3. The statement of work generally acceptable to DARPA for contract negotiations and SDC Proposal No. 79-5719, dated 7 March 1979, have been forwarded direct to DSS-W.
4. The provisions of this Order are in accordance with Section 601 of the Economy Act of 1962 as amended (31 U.S.C. 686).
5. This Order makes available \$75,000 under appropriation and account symbol "9790400.1311 3759 P9P10 1229 S49447 9 DPAA 3759 Research, Development, Test and Evaluation, Defense Agencies," for obligation on behalf of DARPA, only for the purposes and in the amounts specified herein. These funds are not for the construction of facilities.
6. The DARPA/Agent General Requirements, as stated in DARPA Instruction No. 21, dated 13 February 1978, are included herein by reference.
7. Request reimbursement billings to CIA in accordance with paragraph 2 of referenced MIPR.

David Braunstein
David R. Braunstein
Director
Program Management Office

Attachment
CIA MIPR 900907



STANFORD RESEARCH INSTITUTE
MENLO PARK, CALIFORNIA 94025
(415) 326-6200

August 22, 1977

Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attention: Vinton G. Cerf

Gentlemen:

We are submitting herewith for your consideration SRI International's Proposal No. ISU 77-117 entitled "Packet Radio Speech Experimental Program." Also enclosed are executed DD Forms 633-4.

If you desire further information of a technical nature, please do not hesitate to contact Dr. Earl J. Craighill. Contractual matters should be directed to the attention of the undersigned.

Very truly yours,

Spencer Flood
Sr. Contract Administrator

SF/jh

Enclosures



STANFORD RESEARCH INSTITUTE
MENLO PARK, CALIFORNIA 94025
(415) 326-6200

August 22, 1977

Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attention: Vinton G. Cerf

Gentlemen:

We are submitting herewith for your consideration SRI International's Proposal No. ISU 77-117 entitled "Packet Radio Speech Experimental Program." Also enclosed are executed DD Forms 633-4.

If you desire further information of a technical nature, please do not hesitate to contact Dr. Earl J. Craighill. Contractual matters should be directed to the attention of the undersigned.

Very truly yours,

Spencer Flood
Sr. Contract Administrator

SF/jh

Enclosures



STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 · U.S.A.

10 August 1977

Proposal for Research

SRI No. ISU 77-117

PACKET RADIO SPEECH EXPERIMENTAL PROGRAM

Part One--Technical Proposal

Prepared for:

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attn: Vinton G. Cerf

Prepared by:

Earl J. Craighill
Research Engineer
Telecommunications Sciences Center

Approved:

Donald L. Nielson, Director
Telecommunications Sciences Center

Earle D. Jones, Executive Director
Information Science and Engineering Division

CONTENTS

I	SUMMARY		1
II	BACKGROUND		4
	A. SRI Participation In Related Areas		4
	B. Packet Radio Speech Implementation		5
	C. Packet Speech In Other Networks		8
III	OBJECTIVES		10
IV	METHOD OF APPROACH		11
	A. PRnet Testing and Evaluation Using Voice Traffic		11
	B. Expanded Internet Services For Voice Traffic		12
	C. Conference CONTROL Specification and Implementation		13
	D. CVSD Hardware Construction		14
V	HARDWARE AND FACILITIES		16
VI	WORK STATEMENT		17
VII	ORGANIZATION AND SCHEDULING		18
APPENDICES			
A	ROBUST CONFERENCE CONTROL SPECIFICATION		1
B	NETWORK RESOURCE REQUESTS		1
	A. Types of Conferences		1

B.	Tailored Connections	2
C.	User Negotiation of Speech Connections	3
D.	Network Resource Request Summary	5
C	EXTENSIONS TO ROBUST CONFERENCING	1
A.	Unstructured Conferencing	1
B.	Broadcast Packets	2
C.	Limited Feedback Conferencing	2
D.	All-to-All Conferencing	3
D	INTELLIGIBLE INTERRUPTION TECHNIQUES	1

REFERENCES

I SUMMARY

This proposal presents a 12-month program for continued research in digitized packet speech communication systems. SRI International participated in the initial ARPAnet speech transmission experiments and implemented a voice communication capability on the PRnet during the current year under contract N00039-76-C-0364. This experience leads us to propose research in four areas:

- (1) Continued experimentation and testing to assess the impact of wideband (CVSD) speech on the PRnet.
- (2) Development of concepts whereby multinetwork capabilities can be matched to digital speech transmission requirements.
- (3) Development and testing of robust speech conference control protocols.
- (4) Continued development of speech activity detection for use in asynchronous communication systems.

The motivation for this research is to determine the capacity of the PRnet to support limited speech traffic, in addition to data traffic, in a tactical environment where mobile radios are used for conversations with several other individuals, possibly on other networks. This should be accomplished without a severe degradation of the primary network traffic (say, intercomputer logistic information). The PRnet experimentation and network capabilities study needs to be performed in a timely fashion to support the ongoing development of PRnet and internet elements, such as an end-to-end (ETE) transport protocol (TCP or its successors). The implementation and testing of robust conference control protocols and speech activity detection development tasks are costed as options in this proposal. It is recommended that these options be funded at the same time, because of their potential effect on the development of critical network elements and the need for more experience with and testing of realistic conferencing systems.

The continued ETE testing of the PRnet, using simulated and actual speech conversations, is an extremely sensitive measure of network performance and an excellent benchmark against which to compare new releases of hardware and software. In addition to sufficient throughput, digital speech transmission requires a bound on the maximum

delay for each packet so as to preserve speech quality. Any delay-inducing anomaly may be discernible, since speech quality degrades proportionally as the number of lost (or excessively delayed) packets increases. This stresses the importance of using actual speech conversations to evaluate the performance of such new network features as mobile handoff and point-to-point routing.

Internet packet speech communication is a new activity; hence several network elements have not yet been optimized to provide the required service. In particular, speech traffic will affect, and be affected by, the characteristics of the internet transport protocol (TCP or its successors) and the individual networks (such as the PRnet). Packet speech conferencing is one example of a generic service requiring real-time (low delay) network response and multidestination routings. It therefore serves as a concrete example of an internet service requiring capabilities other than the ETE reliable internet protocol thus far developed (e.g., the TCP internet protocol). Selection of more general types of internet service will be influenced by the capabilities of various different networks to accommodate these new services. The proposed PRnet testing program will allow direct assessment of some real-time types of service (TOS). By TOS, we mean the ability of a network user to specify (or negotiate) a minimum set of requirements tailored to his desired communication, say a throughput of 16 kbits/s full-duplex with a maximum end-to-end delay of 300 ms. Mechanisms must be developed, then to dynamically allocate the finite network resources to satisfy the possibly, diverse needs of all users.

Voice conferences are commonly required in tactical communication. These conferences frequently are conducted under stressed circumstances where one, or possibly more, elements or nodes of the communication system may fail. It is undesirable to have such component failures result in a catastrophic loss of communication. SRI has distributed a preliminary specification of a conference control protocol with features of robustness and adaptability to varying network performance that also allows more natural interchanges between conference participants. Consequently, we propose to further develop robust conferencing systems that will provide the desired controlled conferencing capability, and include reliable recovery mechanisms to ensure graceful degradation as a result of element failure. Because of the increased number and variety of packet networks, emphasis has been placed on modularizing the protocol to achieve as much network independence as possible. The PRnet could be used to test an implementation of the protocols (using a specific network interface module) and evaluate the performance and user effectiveness of the system.

Packet communication networks have an inherent capability for providing controlled conferencing. Such facilities allow users to share a common transmission channel and hence packet broadcast addressing is feasible. In addition, the processing capability at each network node allows more sophisticated control and handling of speech conference traffic.

Full-duplex speech transmission traditionally has required that voice data be transmitted continuously at full bandwidth, regardless of the usable speech content. Only when bandwidth requirements were very critical, such as for the Atlantic submarine cable, were some type of speech detection methods attempted. The TASI concept was developed to effectively create half-duplex voice communication, but the crudeness of the approach caused some loss of speech content as well as poor user effectiveness (loss of interruption capability, for instance). The advent of digitized voice service and resulting demands for more bandwidth has renewed interest in good methods of speech activity determination. Fixed assignment channels (circuits) still cannot use such methods and hence are idle 50 to 70% of the time. Asynchronous channels, such as packet systems, can exploit the speech duty factor and still maintain full-duplex connectivity. Speech-activity-regulated transmission requires a reliable method of speech activity detection (SAD). To be effective, any SAD mechanism must be capable of removing impulse noises as well as constant-level background interference without losing normal speech transients or low-level sibilants. This may be achieved either by software analysis on the encoded speech stream, or by a special SAD device, as proposed herein.

In previous and current contracts with ARPA, SRI has participated in the ARPAnet speech conference demonstrations and implemented protocols for voice traffic on the PRnet. In Section II we discuss the details of that experience relevant to the proposed effort. In Section III we define the objectives of the proposed research. In Section IV we discuss the proposed method of approach, including a description of a mechanism for users to request network resources, a description of the basic robust conference control protocol, and some extensions to allow more natural conferencing. In Section V we discuss hardware and facilities; in Section VI we present the work statement; and in Section VII we discuss organization and staffing.

II BACKGROUND

SRI has been actively engaged in voice communication research for some time. In particular, SRI has participated in the ARPA network speech compression (NSC) program for four years. Our focus has been on developing voice communication systems with improved voice quality and system reliability while demonstrating their feasibility and usefulness. The best method for testing complicated voice communication systems is to perform experiments under actual operating conditions so that users can have hands-on experience, and their comments and critiques can be systematically evaluated. To that end, SRI has implemented protocols to handle voice traffic on the ARPA supported packet radio network (PRnet) testbed.

A. SRI Participation In Related Areas

The advent of large computer-based communication networks (like the ARPAnet) that use advanced digital data transmission methods has increased interest in common voice and data networks. We have been most concerned with improving the quality and reliability of integrated voice and data communication and conferencing systems using currently available hardware. SRI has participated in the initial demonstrations of high quality digital speech transmission at 1 to 2 kbits/s (average) on the ARPAnet. SRI has also demonstrated wideband (CVSD) speech communication over the PRnet during the current contract. The goal of the latter demonstration was to assess the effect of wideband (CVSD) speech traffic on PRnet performance.

SRI has also participated in a voice satellite communication study for DCA of several demand-assignment-multiple-access techniques for the Defense Satellite Communication System of the 1980 to 1990 era. One outcome of that study is the specification of different categories of voice communication, and the significant system design parameters that are necessary to support those categories.

The development of internet packet voice techniques must consider networks with different capabilities. The ARPAnet consists of a number of larger scale computer systems with information data bases connected by land lines. The SATnet provides communication between geographically separated places via leased satellites. The PRnet provides low-cost mobile communications to a local area via radio links.

The different capabilities of these networks and the variety of services available within each network suggest that a range of voice

quality is available. Since the needs of voice users are also varied, one would like to provide the required service via network resource requests to achieve a close correspondence between the user requirements and network resource allocation. In this way, the network service level granted to a user could be a function of the type and priority of his communication.

The PRnet testbed at SRI is an excellent place to assess the capacity and quality of voice communication supported by data networks. The existing nucleus at SRI of operational hardware and software plus experienced personnel is well-suited to performing tests and experiments, as well as giving demonstrations of packet systems. The current year of the SRI packet radio (PR) speech project has focused on addition of speech terminals to the Bay Area PRnet testbed. We intend to use the PRnet speech capability next year for continued testing and evaluation of the PRnet for voice traffic, development of guidelines for type of service (TOS) specifications, and for development of a robust conference control protocol.

It is, of course, a realistic goal to be able to communicate from a point in one net to a point in any other, but we shall begin by considering the PRnet. Our experience with the ARPAnet conferencing indicates that improved tactical conferencing requires system reliability, quick recovery from component failures (both hardware and software), and general robustness. We propose to study the implications and requirements of robust conferencing both in the PRnet and in internet environments during the next year.

B. Packet Radio Speech Implementation

During the current contract year SRI has implemented a speech communication capability on the PRnet. Thus far, the requisite hardware and software for CVSD speech transmission have been specified and implemented, and the salient network requirements for speech traffic have been determined.

1. Available Hardware

The PRnet has three types of components: a station (the central control element), repeaters (radio store-and-forward units), and terminals. The PR speech implementation has concentrated on the end-user portion of the terminals called speech interface units (SIUs). The SIU is hooked to a experimental packet radio unit (EPR) and contains all user interfacing, packet forming, and end-to-end protocol functions. Two of the SIUs are dual-processor LSI-11s equipped with interfaces for the PRnet, CVSD voice digitization, and a user input/display. The hardware configuration is arranged so that the high-throughput path of CVSD to/from the PRnet is within one processor while the user interface

and the major control module is within the other. Similar interfaces exist on the NSC PDP-11/40 so that it can be used as another SIU. The PDP-11/40 SIU has, in addition, more memory, two CVSD units, two 1.2 megaword disks, and a high speed signal processor (SPS-41).

2. Completed Software

The network voice protocol (NVP) specifies two parts: the VOICE protocol to handle the high-data-rate speech data; and the CONTROL protocol to establish and maintain connections, determine vocoder parameters, and supply status information to the user. For the PRnet, SRI has implemented a specific VOICE protocol compatible with the NVP specification for the ARPAnet. The software is written to run under a reduced ELF operating system called MOS (micro-operating system) that uses very little memory and is optimized for high throughput. MOS provides multiple process scheduling (round-robin), timer control, and I/O queueing. The VOICE protocol implementation is a speech transceiver with the following properties:

- (1) The speech transceiver is one reentrant process, with functionally independent transmitter and receiver modules. It makes no assumptions about the hop transport protocol, and is independent of SIU configuration and speech device hardware (all speech hardware dependent functions are performed by a device driver package). There are also provisions for multiple extensions for each SIU and conditional inclusion of statistics collection code.
- (2) The receiver uses packet sequence number windowing to achieve ordering, duplicate filtering, and lost or late packet detection, with automatic receive window resynchronization when necessary. Lost packet substitution is performed, but can be preempted in case of late packet arrival. Reception from any of several connections is selectable at the direction of a local or foreign conference CONTROL program.
- (3) The transmitter performs speech activity determination via a finite state machine algorithm, with transmitter shutdown during silence. Simultaneous multiconnection transmission (with provisions for packet broadcasting) occurs at the direction of a local or foreign conference CONTROL program. Provisions have been made for dynamic speech data flow-control.
- (4) The speech device driver contains all device-dependent code and interprets device-independent protocol parameters (such as interpacket interval). The driver

returns a device-independent tag with each packet that quantifies its speech content for use in speech activity determination.

- (5) Speech activity detection (SAD) is accomplished by using the device-independent tag from the device driver as well as the current state to generate a next state ("speech", "silence", "coming out of silence", or "going into silence") and an action (requesting or inhibiting new vocoder input, transmitting or discarding the new packet). In this manner the packet type and trend information is used with the speech activity history to govern transmission of speech packets.

As part of the current contract, it was proposed to implement the CONTROL portion of the network voice conference protocol (NVCP) in the SIUS on the PRnet. A careful review of NVCP indicates that extensions can be made that will better match it to the capabilities and characteristics of the PRnet and future internet conferencing. They include:

- . Allowing the use of broadcast packet addressing (when available)
- . Using an initial connection protocol compatible with TCP
- . Allowing for conference partitioning into subconferences and subsequent recombining (such a condition could be forced, by loss of a GATEWAY between two networks for example).

3. PRnet Performance Evaluation

During the current contract, end-to-end (ETE) performance evaluation of the PRnet was initiated to determine the effect of the PRnet hop transport protocol on speech communication performance. The standard tests that measure reliability (number of lost or duplicate packets), and bandwidth (number of packets per second) were expanded to determine the ETE delay characteristics. Speech transport techniques require detailed knowledge of the delay distributions (not just the average value) to preserve speech quality. This effort was especially valuable because the effects of placing high data-rate traffic on the PRnet were also observed.

To accomplish this, the speech transceiver software was modified for use as a traffic source in ETE measurements, enabling generation of simulated speech traffic with precisely controllable parameters. In addition to the usual performance measurements of throughput, reliability (lost and duplicate packets), and average delay,

histograms and time-histories of delay needed to be determined. For this reason, a more sophisticated ETE measurement facility was developed that permitted precise, centralized control of all network elements (traffic sources and EPRs), as well as ETE timing synchronization for delay measurements. More minute details of the channel access protocol (CAP) were thereby exposed, including retransmission times, bidirectional traffic blockages, and other features that affect speech ETE delay. These measurements are extremely valuable in analyzing CAP performance and uncovering features and capabilities of new releases of CAP.

C. Packet Speech In Other Networks

In addition to the NVCP protocol implemented for the ARPAnet [1], another conference control protocol is being developed for the ARPA SATnet [2]. While adequate for their specific environments, these protocols rely on unique network features. The design of a conference protocol should take these differences into account, but modularize the network-dependent aspects to allow future extensions to internet conferencing.

The ARPAnet is characterized by distributed store-and-forward transmission over land lines with pipelining and some noncontention service; the distributed transmission resources can be used simultaneously by several users if they are separated geographically (unlike for the satellite case). The delays are moderate and principally caused by contention for network resources, but grow for participants at remote or busy portions of the network. A natural broadcast addressing mechanism does not exist.

The SATnet being developed by ARPA will be characterized by a highly efficient demand assignment transmission, but with delays that are multiples of the satellite propagation delay (270 ms). The satellite transponder provides a natural broadcast channel so that one packet can be received by all users almost simultaneously. The satellite resources will be allocated through the stream contention-priority-ordered-demand-access (C-PODA) algorithm, so a speech conference need share its channel only among its participants. A satellite network has the potential for providing a very high-capacity link, although the initial experiments will be performed with a 64-kbits/s channel.

The PRnet is characterized by relatively short delays (caused principally by retransmissions) and high capacities. The retransmissions result from either contention or failure in acquiring the rapid bit synchronization that is required for the spread-spectrum receivers. The PRnet has a natural broadcast capability although the delay is not as uniform, because of multiple repeater links, as it is in the SATnet.

The PRnet offers the possibility of exploiting carrier sense to increase the effective channel capacity. However, some terminals may be hidden and hence it is not always possible to realize this improvement. An advantage can be realized, however, since all users do not necessarily compete for the same resources of bandwidth and power, as they must in the SATnet. Therefore, users at one edge of the PRnet might not interfere (through resource contention) with users at the opposite edge.

The PRnet now lacks an effective method for reserving channel capacity for a conference, thus interaction with other network users will have to receive greater attention than in the case of the satellite network.

III OBJECTIVES

The objectives of this research are to assess the effect of wideband (CVSD) speech on the PRnet, and to develop concepts whereby multinetwork capabilities can be matched to requirements for digital speech transmission. Objectives of the optional part of this research are to develop a robust packet speech conference control protocol that maintains compatibility with internet requirements, to test the effects of component failures and other network dynamics on voice conferencing performances, and to develop speech activity detection hardware.

IV METHOD OF APPROACH

SRI has been implementing a speech capability on the packet radio network during the current contract. The PRnet testbed at SRI is an excellent place to test packet voice communication networks. The existing nucleus of operational hardware and software plus experienced personnel is well-suited to performing tests and experiments, as well as giving demonstrations of packet systems. The current year of the SRI PR speech project has focused on addition of speech terminals to the Bay Area PRnet testbed. We intend to use the PRnet speech capability in the following year for continued testing and evaluation of the PRnet for voice traffic, development of guidelines for type of service (TOS) specifications, development of a robust conference control protocol, and for further development of CVSD hardware with speech activity detection.

A. PRnet Testing and Evaluation Using Voice Traffic

We propose to continue the ETE performance evaluation as new PR software becomes available. The usual performance measurements concentrate on reliable throughput. For speech, too much delay may be incurred for totally reliable delivery, and hence there is a more stringent requirement for packet voice communication, that is, the preservation of voice quality. The requirements of speech traffic, particularly the bound on maximum ETE delay, are good benchmarks for the PRnet. In addition, transient network glitches that are not normally recorded in averaged measurements are immediately apparent during a voice conversation. We thus propose to use a network configuration with at least one mobile SIU in order to adequately exercise the software that we expect will be released during the next year. Some of the specific features of the new PR software that will be pertinent to voice traffic are:

- . Assignment of point-to-point (PTP) routes, and
- . Mobile handoff.

As an example of the effect of voice traffic requirements on PRnet protocol design, we will discuss a proposed method for PTP routing and mobile handoff. The PTP routing proposed by Sussman [3] attempts to select a minimal delay path, which need not include the station. This route may include several hops, but does circumvent the station bottleneck and thus is more suitable for speech traffic. Currently, it is proposed that all routes be assigned by the station [3]. The

control strategy, although simplified, is vulnerable to high link delays because of large nets, equipment failure, or traffic congestion. When a mobile terminal moves out of its assigned repeater's range, and a speech packet fails on the primary route, the round-trip time required to get an alternate route from the station might introduce so much delay that the terminal-receiver would discard the packet anyway.

This problem can be partially alleviated by keeping some local connectivity information in the EPRs. Sunlin [7] has proposed collecting "good neighbor" information by issuing local repeater on packets (ROPs) and listening for an acknowledgement. This information can allow for selection of locally optimal alternative routes whenever the station control link is unavailable or too slow. We would recommend a strategy where the station sends out primary and secondary routes when a PTP route is requested. Then the EPRs will use this information along with their local connectivity information to keep a packet moving through the net on the best possible route without introducing long delays. Since a route is expected to remain valid for a time longer than it takes to relabel, the station should be advised of any alternate routes that are required so that a new globally optimal route can be assigned for the connection (note that if the assumption about route life-time is true, or if route changes amount to bouncing back and forth between primary and secondary routes, then two routes are sufficient). These strategies better meet speech transmission requirements because service gracefully degrades (small delays are added) rather than being interrupted for discernible periods.

Speech traffic will be the most telling test of these situations. We propose to add a voice link to be used over the PRnet with at least one mobile terminal. The voice conversations can be closely monitored for glitches and the packet and route change statistics can be examined to determine the cause and extent of voice degradation.

B. Expanded Internet Services For Voice Traffic

The design of packet communication systems has been guided strongly by the requirements of data rather than voice traffic. In expanding the traffic load to include voice and others such as real-time video, it will be exceedingly difficult to design a network with fixed capabilities suitable for all varieties of traffic. Indeed, it may be more economical to adapt existing capabilities to traffic requirements by specifying a type of service (TOS). The testing of the PRnet with speech traffic, for instance, might yield a set of hop transport protocol parameters that are optimal for wideband speech in the PRnet, but are different than the set selected for standard data traffic. If one then defined a TOS label for speech packets, these parameters would be used to transport speech packets on that connection. Of course, this procedure can be made to allow several TOSs for speech (casual conversations, important discussions, emergency communications, and so

on), and for data traffic. The TOS requests must be integrated with the particular ETE protocol. In the speech case, we propose to investigate the relationship of TOS with the conference control specification that was developed under the current contract.

C. Conference CONTROL Specification and Implementation

The development of robust conferencing protocols will have three phases. The first phase (Task 2) will be a review and update of the robust conference CONTROL protocol specification that was drafted during the current contract (Appendix A gives an overview of the specification). Ways to match user needs with the capabilities of packet systems will be studied, with emphasis on identifying the properties of the underlying networks that affect the conferencing performance. One specific mechanism for supporting network resource requests will be analyzed in depth (see Appendix B for more detail).

The second phase (Task 3) is costed as an option in this proposal and will be the software implementation of the robust conference CONTROL protocol and integration with the existing VOICE transceiver. This phase is necessary to confirm the general design and uncover limitations, and during it SRI will remain cognizant of future packet radio improvements. The phase will start at ARPA's discretion and after the review and updating in Phase One are completed. Performance analyses of the PRnet indicate that there is sufficient throughput capacity to allow a four party conference (two extensions on the PDP-11/40 SIU) for CONTROL protocol check-out. The proposed extensions to robust conferencing that are discussed in Appendix C will be critically evaluated with respect to utility and practicality and implemented as appropriate.

The third phase (Task 3) will include a demonstration of a live conference that will test the integrated CONTROL and VOICE protocol. Specific tests of protocol robustness will be performed by introducing artificial errors and component failures. Measurements will also be taken to evaluate the network performance with conference voice traffic.

An important part of the three-phase protocol development and testing effort is participation in the discussions of internet protocols and experiments. By representing speech users at appropriate meetings, we will also ensure the development of packet technology well matched to speech requirements.

D. CVSD Hardware Construction

Speech-activity-regulated transmission requires a reliable method of speech activity detection (SAD). To be effective, any SAD mechanism must be capable of removing impulse noises as well as constant-level background interference without losing normal speech transients or low-level sibilants. This may be achieved either by software analysis on the encoded speech stream, or by a special SAD device, as proposed herein (and costed separately).

As part of the current contract, we have designed, implemented, and tested a software SAD algorithm that works on CVSD-encoded data. Even without fine-tuning of parameters, this approach has worked very well in tests on PRnet hardware. In brief, we use a CVSD vocoder specially modified to generate an interrupt whenever the number of "010" or "101" patterns exceeds a fixed threshold (an alternating 1-0 pattern indicates silence). Proper adjustment of this threshold maintains the input speech levels in a range optimally suited to the CVSD encoding. Packet read-in is initiated by this interrupt, and continues until the highest level of the SAD mechanism (a finite state machine) shuts off input to await another interrupt.

As a packet is read, the vocoded data are examined, and several statistics are collected. When an entire packet has been filled, a speech activity function is evaluated using these statistics and several programmable constants. The function's range encompasses three logical values that serve to "type" the packet when it is passed on to the SAD state machine. These values, or packet "tags", qualitatively identify both the potential speech content of a packet and the trend (towards or away from silence) that it is taking; the tags are called "silence", "transition", and "nonsilence".

Packets are passed to a higher-level finite state machine that endows the overall SAD mechanism with the ability to take account of speech activity history. It is here that the decisions controlling the transmission of packets and the activity of the input vocoder are made. The operations of this mechanism are independent of the voice digitization method. As currently defined, the SAD machine takes as input the "tag" returned with the new packet, as well as its current state. States include "speech", "silence", and one or more each of "coming out of silence" and "going into silence". Actions that it evokes include requesting/inhibiting new vocoder input, transmitting/discarding the new packet, and so on. The SAD state machine also has the ability to enqueue questionable packets until it has more clearly established a speech activity trend; such packets may then be transmitted or discarded.

Although this algorithm can work with any vocoding for which the packet tagging function can be evaluated, its operation with CVSD vocoding is particularly efficient because of the nature of CVSD

encoding: any segment of the encoded stream may be defined as "silence" if it is comprised of alternating ones and zeroes, and "nonsilence" otherwise. Therefore, the statistics that need to be collected are simple counts.

We propose to design and construct a hardware SAD device based upon our SAD mechanism and using internal CVSD encoding. The device would accept as input requests for new packets (whose attributes such as length and CVSD rate would be programmable), and would generate as output an interrupt when the current vocoder input represents usable speech. A CVSD-encoded packet would accompany each interrupt; the packet could be ignored, used directly, or perhaps utilized in conjunction with a parallel vocoding of the speech stream for imbedding applications. Note that although the SAD circuit would perform CVSD encoding, any other kind of vocoding may be performed in parallel; the output of the SAD circuit would simply be used to regulate packet transmission.

The recent introduction of true single-chip microcomputer systems at low prices leads us to propose implementation of the SAD circuit using an Intel 8085 microcomputer with the SAD algorithm residing in the on-chip PROM. This approach would provide for unlimited flexibility, since SAD parameters would be settable according to specific user needs. The recently introduced CVSD chips will be used to give a full-duplex voice channel. The digitization rates will be independently controllable for transmitter (input) and receiver (output). The majority of functions will be designed around an 8-bit tri-state bus with an interface to the LSI-11 16-bit Q-bus. DMA operation will be considered to reduce the LSI-11 CPU overhead.

This prototype design will be very useful in investigating the hardware implementation aspects for CVSD voice digitization. Whereas many LSI chips will be used to reduce space and power consumption, the packaging will allow modification as needed for such initial engineering designs. The demonstrated performance of the SAD algorithm indicates that it will be a very useful addition to a voice digitization terminal that must operate in very noisy tactical environments.

V HARDWARE AND FACILITIES

The current PR speech contract has focused on implementation of a three-node PR-speech-net. Two LSI-11-based SIUs are operational and the NSC PDP-11/40 is a third with some additional capabilities: two CVSD extensions and a disk storage system for data or measurement logging. The PDP 11/40 is also used for down-line loading of the LSI-11 SIUs. Existing maintenance contracts for these machines should be extended to give continuous coverage.

As an option, a third LSI-11 SIU should be purchased in order to allow one four-party conference (with each user at a unique EPR site) or two simultaneous conversations.

Another hardware activity proposed for next year is the construction of miniaturized CVSD units with programmable controls (rate and packet length) and SAD hardware as described in the method of approach, Section IV. We propose building at least four such units.

VI WORK STATEMENT

TASK 1: Carry out the ETE speech measurement and testing of the PRnet for new releases of hop transport protocols (CAP) to ensure performance and reveal the effect of PRnet changes on speech transmission; implement a voice link on the PRnet with at least one mobile SIU; characterize and document the impact of voice traffic in the PRnet.

TASK 2: Review and update the specification of packet voice conferencing CONTROL protocols that was drafted under contract N00039-76-C-0364 with respect to robustness, user requirements, and network structure and control; develop the language and structure of the user/network-overseer negotiation for network resources with emphasis on providing network support elements for conference CONTROL protocols in an internet speech environment.

TASK 3: Design and implement the programs for the robust conference CONTROL protocols specified under Task 2; test the robustness of the speech conferencing system on the PRnet (costed as an option). Acquire and integrate a third LSI-11 Speech Interface Unit (costed separately).

TASK 4: Design and building four CVSD input devices with programmable controls and speech activity detection hardware (costed as an option).

VII ORGANIZATION AND SCHEDULING

The proposed effort will be conducted entirely within the Information Sciences and Engineering Division of SRI. Most of the personnel who will be working on this contract are in the Telecommunications Sciences Center. Close coordination will be maintained with the Packet Radio Implementation Project.

Task 1 will be performed as needed when new software becomes available for testing in the PRnet.

Task 2, the review and integration of the conference CONTROL specification, will be started at the beginning of the contract period. When all comments have been integrated and the necessary network resources and capabilities are available, the conference CONTROL protocol can be implemented and integrated with the VOICE protocol (Task 3). The integrated protocols can then be tested on the PRnet for robustness and user effectiveness. Since the length of time necessary for review of the conference control specification is not known at this time, Task 3 has been costed as an option to be funded at ARPA's discretion.

Task 4, the design and build of four CVSD units, will begin at contract award time, if ARPA elects to fund it. Thus, the hardware can be used for the tests and experiments.

Biographies of principal SRI personnel who will work on the project follow.

DONALD L. NIELSON, DIRECTOR
TELECOMMUNICATIONS SCIENCES CENTER
INFORMATION SCIENCE AND ENGINEERING DIVISION

Specialized professional competence

- . Application of communications technology; radio propagation; radio channel characteriziation; packet-switched digital networks; ionospheric physics; nuclear effects on communication systems

Representative research assignments at SRI (since 1959)

- . Supervision of implementation and operation of packet radio network
- . UHF radio channel measurements and characterization in urban areas
- . Study of packet radio systems
- . Ignition noise measurement and suppression studies
- . Radiation of nonsinusoidal signals
- . Application of optics to signal processing
- . Investigation of long-range transequatorial propagation at VHF
- . Development of techniques and models for predicting the effects of nuclear explosions on HF communication circuits
- . Investigation of the application of oblique-incidence sounders to HF communication
- . Study of problems in HF and VHF propagation, including ground backscatter, nuclear effects, prediction, and the use of ionospheric radars

Other professional experience

- . U.S. Army Security Agency: concerned primarily with radio propagation and communication problems

Academic background

- . B.S.E.E. (1956), University of Idaho; M.S.E.E. (1962) and Ph.D. in electrical engineering (1969), Stanford University

Publications

- . Author or coauthor of more than 20 SRI technical reports in the above research areas
- . Author or coauthor of more than 12 technical papers presented at national and international symposia and published in proceedings or technical journals

Professional associations

- . Registered Professional Engineer, California
- . Institute of Electrical and Electronics Engineers
- . U.S. URSI Commission E

D. THOMAS MAGILL, PROGRAM MANAGER
TELECOMMUNICATIONS SCIENCES CENTER
INFORMATION SCIENCE AND ENGINEERING DIVISION

Specialized professional competence

- Satellite communication systems; modulation-demodulation theory; statistical estimation theory; tracking system theory; voice digitization

Representative research assignments at SRI (since 1970)

- Project supervisor for DCA project to develop a computer program to optimally distribute satellite power and bandwidth to frequency-division users
- Project leader for ARPA effort developing a real-time linear predictive vocoder
- Project leader for effort to find a more efficient method of encoding the speech excitation signal
- Comparison of multiple-access techniques suitable for DSCS-Phase II
- Signal design for HF band tactical radio sets
- Statistical analysis of the usage of mobile radio bands

Other professional experience

- Department manager/staff scientist, Communications Systems Department, Space and Reentry Systems Division, Philco-Ford Corporation
- Developed a unified theory for the extrinsic noise performance of tracking cross correlators
- Investigated multiple- and random-access communication systems, world-wide timing systems, and target location systems
- Conducted some of the first time-division multiple-access experiments with a hard-limiting satellite transponder
- Analyzed, designed, and directed the development of high-rate quadriphase modems
- Directed the development of two experimental ground terminals for a TDMA satellite system
- Instructed graduate courses in communication theory

Academic background

- B.S. in electrical engineering (1957), Princeton University; M.S. (1960) and Ph.D. (1964) in electrical engineering, Stanford University

Professional associations and honors

- Institute of Electrical and Electronics Engineers (Groups on Acoustics, Speech, and Signal Processing, Communications, Control Systems, and Information Theory; chairman of Information Theory Group, San Francisco Section, 1970-71; group chapter coordinator of San Francisco Section, 1970-73; chairman of Control System Society Chapter, Santa Clara Valley Section, 1975-76)
- Technical Program for WESCON (cochairman, 1973)
- B. H. Whiting Memorial Scholarship (1953-57)
- Lockheed Honors Cooperative Program (1961-63)

EARL J. CRAIGHILL, RESEARCH ENGINEER
TELECOMMUNICATIONS SCIENCES CENTER
INFORMATION SCIENCE AND ENGINEERING DIVISION

Specialized professional competence

- . Artificial intelligence; pattern recognition; applied mathematics; communication theory; computer and communication networks; time series analysis; man-machine interfaces; real-time signal processing; interactive display system design; talker identification; spoken word recognition; speech compression

Representative research assignments at SRI (since 1968)

- . Designed and implemented interactive graphic display program
- . Developed acoustical processing for automatic speech-recognition system including formant tracking, digital filters, and frequency analysis
- . Developed speech bandwidth compression algorithms
- . Designed and implemented low-bit-rate speech transmission system on ARPA Network
- . Programmed a high-speed parallel signal processor for real-time operation
- . Developed protocols for asynchronous demand-access speech communication network
- . Designed speech teleconferencing system
- . Studied the system requirements for voice transmission over a satellite communication system

Other professional experience

- . Electronics Research Laboratory, Montana State University: developed computer programs for displaying multidimensional data and analyzing classification procedures; taught applied mathematics
- . Michigan State University: developed and simulated a functional model of mammalian brain structures using nonlinear decision algorithms; applied these computerized decision algorithms to speech recognition

Academic background

- . A.B. in mathematics and physics (1963), Carroll College; M.S. in electrical engineering (1965), Montana State University; Ph.D. in electrical engineering (1971), Michigan State University

Publications

- . Coauthor of description of function model of the reticular formation, in *The Mind: Biological Approaches to Its Functions*, W. C. Corning and M. W. Balaban, eds. (1968); "Automatic Speech Recognition Based on a New Segmentation Procedure," Ph.D. thesis (August 1971); coauthor of "Human Rod Visual Delay," *Nature* (January 1968)

Honors

- . Sigma Xi

V. DONALD CONE, RESEARCH ENGINEER
TELECOMMUNICATIONS SCIENCES CENTER
INFORMATION SCIENCE AND ENGINEERING DIVISION

Specialized professional competence

- Radio propagation; communication instrumentation and measurements; antenna design; design of protective shields and instrument shelters; radio/digital systems

Representative research assignments at SRI (since 1954)

- Single-sideband equipment development
- VLF satellite instrumentation
- Airborne backscatter measurements
- Ionosphere sounder measurements
- Project leader on propagation measurement projects
- Project leader on a radio noise and interference survey
- EMP measurement programs
- Airborne antenna pattern measurements
- Vehicle design for mobile instrument systems
- Packet radio network installation, test, and maintenance

Publications and patents

- Author or coauthor of several reports on ionosphere sounding techniques as related to communication systems
- Patent on display system for real-time use of ionosphere sounder data

KEITH S. KLEMB, COMPUTER APPLICATIONS ANALYST
APPLICATIONS PROGRAMMING GROUP
COMPUTER OPERATIONS AND SERVICES

Specialized professional competence

- Computer programming in FORTRAN, COMPASS, PL1, and 360 ASSEMBLER languages; IBM, CDC, PDP, INTERDATA, and microcomputer experience; hardware-software integration experience; systems orientation; experience with most time-sharing systems; mobile systems experience; interfacing with design engineers

Representative research assignments at SRI (since 1969)

- Application of SPSS and DATATEXT commercial statistical packages
- Maintenance and augmentation of SPSS statistical package at this installation
- Authorship of general program for analysis of telephone network use and cost
- Authorship of software driver for the Benson and Lehner flatbed plotter
- Maintenance and improvement of computerized billing systems
- Analysis of data recorded on a variety of field equipment
- Application of Benson and Lehner data reduction equipment (LARR-V, 29E)
- Design and implementation of large-scale data processing systems
- Provided several projects with high speed utility programs
- Authored graphics package for nuclear defense modeling program (SEER)
- Editing and preprocessing routines for U.S. government weather data tapes (three million records)
- Maintenance and enhancement of operating systems on minicomputers
- Software support for the (ARPA) packet radio project
- Senior data reduction aide, Special Services, Computer Planning and Operations

Academic background

- A.A. in scientific computer programming (1973), De Anza College; minor in advanced mathematics (1965-67), Foothill College

DARRYL E. RUBIN, RESEARCH ENGINEER
TELECOMMUNICATIONS SCIENCES CENTER
INFORMATION SCIENCE AND ENGINEERING DIVISION

Specialized professional competence

- . Computer networking and communication protocols; computer operating system design and systems programming; digital logic and micro-computer hardware design

Professional experience

- . Designed and implemented a generalized interhost file transfer package for use on a packet-switched computer network (ARPANET)
- . Designed and implemented software for a stand-alone portable terminal system for direct connection to packet-switched computer networks (ARPANET and PRnet)
- . Designed and implemented software for real-time packetized speech conferencing over a computer network (PRnet)

Academic background

- . B.S. Biology, Stanford University
- . Currently working on M.S.C.E., Stanford University

Academic honors

- . Phi Beta Kappa

Publications

- . "Telnet under Single-Connection TCP Specification," Stanford University Digital Systems Laboratory, Technical Note #76 (February 1976)

Professional associations

- . Institute of Electrical and Electronics Engineers

Appendix A

ROBUST CONFERENCE CONTROL SPECIFICATION

Appendix A

ROBUST CONFERENCE CONTROL SPECIFICATION

This is a summary of a more complete specification of the conference CONTROL* protocol, which will be called a voice communication supervisor (VCS) due to the generality of its functions. The VCS will reside in an SIU along with one or more of the voice communication transceivers (VCTs) implemented during the current contract; the functions of the VCS are to:

- . Provide an effective user interface
- . Mediate establishment of communication with other individuals or conferences
- . Direct the operation of (local or foreign) VCTs
- . Allow establishment of the conference configurations outlined in Appendix B
- . Provide robust automatic recovery from crashes and error conditions.

The CONTROL network is a logical star: conference control is centralized in one VCS, the designated CHAIRMAN, but may shift to any other VCS when necessary since they are homogeneous. The protocol provides for automation of all control functions, and is fully pipelined to minimize control traffic and delay. Command acknowledgment is optional in case control connections are unreliable. The protocol is designed to take full advantage of broadcast addressing if available, without sacrificing efficiency if it is not. However, no assumptions are made about the addressing or connection management mechanisms actually in use, rendering the protocol reasonably network-independent.

INITIAL SETUP is usually automatic: either one VCS may call the others into the conference, or the other VCSs may call into a single VCS. Negotiation as a part of initial setup allows rapid optimization of the conference configuration. In case of multiple callers trying to establish the same conference, initial CHAIRMAN assignment occurs as part of this negotiation to prevent deadlocks.

* All capitalized words refer to specific program code segments rather than the more general standard meanings.

NEGOTIATION allows the CHAIRMAN to interact with all conferees simultaneously for rapidity, with a special "capabilities" message sent by each conferee to the CHAIRMAN to minimize negotiation traffic. There is flexibility in specifying all negotiable parameters, which include vocoder parameters, assignment of the floor or chair, conference configuration, and the like; renegotiation may occur at any time.

CONTROL OF VOICE STREAMS is pipelined and centralized at the acting CHAIRMAN for efficiency. There are provisions for TALKER handoff by priority, by request, or as a result of inactivity. The VOICE network is logically fully connected at all times to reduce TALKER handoff delay, and to allow intelligible interruption or multiple TALKERS.

MAINTENANCE OF PARTICIPANT LIST occurs by dynamic addition and deletion of participants, either automatically (i.e. if a participant crashes) or by request. Optional renegotiation may be triggered by changes in the participant list to maintain optimal conference configuration.

USER INFORMATION is typically supplied to the user device by the local VCS according to protocol without need for special inter-VCS messages, and includes identification of the current TALKER, notification of floor and chairmanship acquisition or loss, participant list updates, and so on.

CHAIRMAN MONITORING AND REASSIGNMENT occurs automatically in case of CHAIRMAN or network failure; CHAIR handoff may also be requested. Provisions will be made for protocol extensions to allow conferences to be fragmented into a hierarchical structure of subconferences.

Appendix B
NETWORK RESOURCE REQUESTS

Appendix B

NETWORK RESOURCE REQUESTS

Digital communication systems can be better matched to a variety of military speech functions than analog channels, so a digital communication system can potentially offer better service to a diverse user community. However, to achieve the flexibility and realize this potential improvement, it is necessary to interrelate three aspects of the overall communication system.

First, the user population must be characterized from dimensions that include communication or conference TYPE, traffic pattern (duty factor), average talk spurt, and average off period. A set of user categories (denoted TYPE/GRADES) can then be created.

Second, it is necessary to characterize the generic service (denoted GENERIC) requirements of each of the user categories. For example, a user may want to specify the data reliability and delay, the required channel capacity or data rate, and the cost and complexity of this service.

Third, it is necessary to interrelate the generic service requirements with the radio communication system design concept and parameter values (denoted SPECIFIC). For example, in the PRnet a given maximum acceptable delay may indicate that only two retransmissions should be tried at each repeater hop.

A. Types of Conferences

Many different TYPES of conferences can occur in addition to the ones that naturally suggest themselves to research engineers familiar with round-table discussions. In the tactical military environment, the conferences are likely to be more structured with a definite hierarchy; one can imagine a commanding general wishing to confer with the commanding officers of each of the component forces under the general's control. This hierarchical structure raises some difficult issues in conferencing strategy, such as how to grant the floor quickly to a lower echelon officer with high priority information, and how to maintain reliability under stressed circumstances, when communication is most essential.

The following list identifies the major conferencing modes that we expect may occur in tactical environments. Some of these represent degenerate forms of conferencing but are listed because a multiuser

conferencing system should be able to accommodate these modes. The major modes are:

- . Parliamentary conference (central chairman with queued requests)
- . Round-table conference (anyone who talks gets the floor)
- . Hierarchical conference (chairman and subchairman structure)
- . Dispatcher mode conference (most of the traffic flow is radial)
- . Polled conference (chairman periodically selects each talker)
- . Broadcast message (majority of users are listening)
- . Separate two-party, full-duplex communication
- . Hybrid combinations of the above.

Round-table conferences appear well matched to the viewpoint that the current talker is the conference controller. However, this concept of floating conference control is not well suited to hierarchical or dispatch mode conferences that are frequently required in tactical environments. Consequently, it is desirable to develop a conferencing system applicable to all of these conference types.

B. Tailored Connections

Another way to improve speech transmission is to tailor the connection to the voice service desired by the user. To best provide the requested service, each node could utilize transmission parameter values specific to each logical connection. For example, values of packet characteristics, such as packet length and transmission rate, could be used to adjust CAP timing parameters. Currently, these parameters are set to worst-case values that reduce collisions for any condition, but also add unnecessary delay for short packets at 400 kbits/s rates. During the CAP testing, we will investigate the relationship between packet lengths and transmission rates, and the resulting ETE delay parameters.

Not all relevant protocol parameters can be adapted in such a simple manner; CAP may need additional connection information. With this information, GENERIC requirements--such as reliability, bandwidth, and delay--could be related to SPECIFIC protocol implementations. For instance, the optimum number of retransmission attempts is a function of end-user needs. Since six retransmissions can occur at each EPR, resulting in ETE delay beyond reasonable bounds for speech, the number of retransmissions should be limited for speech connections. EPRs

should not be dedicated to each connection, so the extra information needed to set connection-specific parameters should be carried in the packet.

We propose to study these parameters in order to determine their relationship to ETE requirements. We also propose to investigate methods for matching the diverse user needs with specific protocol parameter settings in each EPR.

C. User Negotiation of Speech Connections

Designing a network that is optimal for all users is clearly difficult. We propose to investigate methods whereby users can negotiate with a network master control process (MCP) for voice service that best satisfies both the user's needs for the particular communication he intends and the efficient allocation of network resources. The nature of the protocol required between the user and the MCP is one of the key issues. We have identified three levels of voice service specification and have discussed how a correspondence between the three levels can be established. We propose to use the levels as follows (see Figure B-1): voice TYPE/GRADE level between the user's local control process (UCP), so that nonrealistic requests can be quickly rejected; GENERIC voice level for negotiation between the UCP and the MCP (independent of specific net characteristics); and finally, SPECIFIC voice level for communication with the nodes (EPRs) within each net. The latter portion can be the COMMUNICATION PROFILE suggested by Cohen [4].

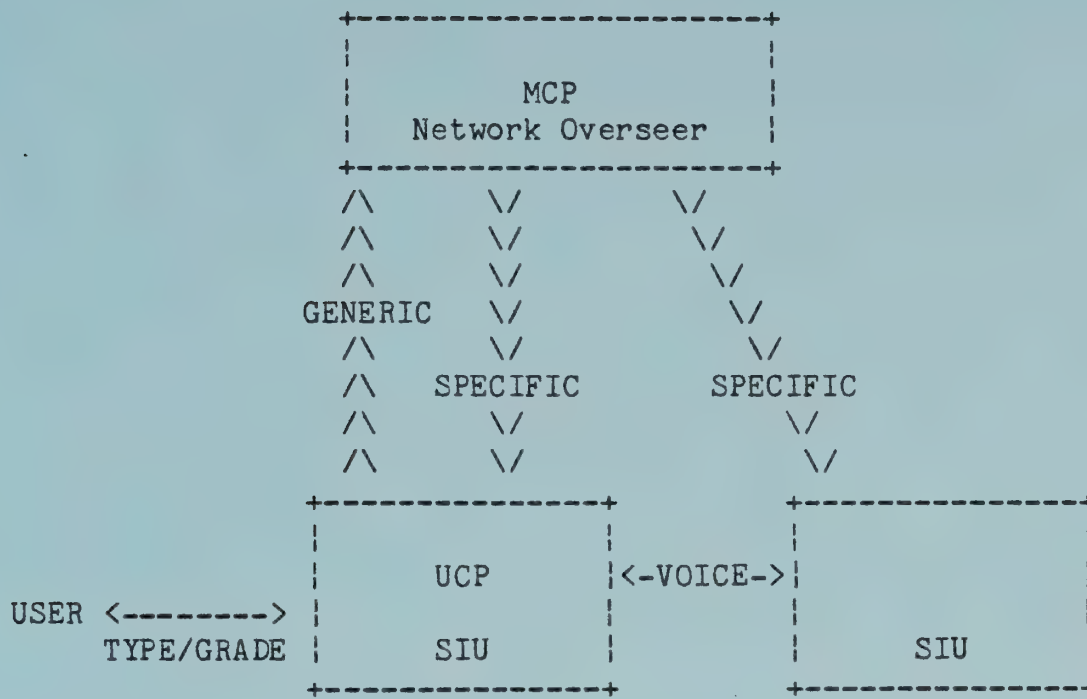


Figure B-1. Protocols For User Setup and Voice Connection

Sussman[5] has specified a method for assigning PTP routes that uses the reliability and delay of each hop to pick an optimal route. This method can be expanded to consider data-rate requests as well as nominal values of delay and reliability. The station control process would then acknowledge the service request with respect to current network traffic, operational elements, and location of the connection-end terminals. Priority and resource cost functions could be included easily in such an interaction.

Another item needing investigation is the accuracy and timeliness of hop reliability and delay estimates. For the best assignment and rapid updating of routes in dynamic network configurations, these estimates must provide quick and accurate hop information. The current information available to the connection process does not identify broken links between nodes very rapidly, so it takes a long time to discover that the station is actually causing bottlenecks by assigning nonexistent links for PTP routes. The proposed mobile speech exercise can be used to evaluate Sussman's suggested measurements in the framework of providing PTP routes for speech connections.

D. Network Resource Request Summary

When the three aspects (TYPE/GRADE, GENERIC and SPECIFIC) of user requests for network resources are interrelated in an analytical fashion, it is possible to determine the overall system performance. It is also possible to investigate radio communication system modifications that lead to improved performance and how the network resources can be allocated to satisfy user requirements. The main concern is that the mechanism for permitting users to request special service does not lead to a fragile system that readily malfunctions with subsystem failure, or system inefficiency because of a large service traffic overhead. The latter problem might arise if each (short) packet contained the channel service request.

Our approach for dynamic allocation of network resources will include the following considerations:

- . The current low (relative to ideal) data-rate capacity of the PRnet for speech traffic results partly from the hierarchical routing and concomitant bottlenecking at the station. We believe that PTP routing will reduce such bottlenecks and allow sufficient bandwidth for speech connections.
- . We propose, however, that when the routes are assigned by a network overseer (MCP), some of the route characteristics should be determined by a negotiation with the user. The issues in this negotiation process are: is the MCP at one place (the station) or distributed? What is the language of the negotiation, between the user and his UCP, between the UCP and the MCP, and between the MCP and ultimately the nodes (EPRs)?

- . Our proposal for the MCP-NODE communication is that each packet should contain its own connection-specific protocol parameters and alternate route information.
- . The characteristics of these routes must change dynamically with modifications in the network configuration resulting from changes in traffic, operational components, and location of (mobile) terminals.
- . The route changes must be quick enough to keep speech packets moving (maybe in a slightly circuitous route) in order to not totally block the speech connection.

Appendix C

EXTENSIONS TO ROBUST CONFERENCING

Appendix C

EXTENSIONS TO ROBUST CONFERENCING

Speech conferencing can require a great deal of throughput, especially when there are many conferees or several simultaneous talkers (as in a natural conference). Point-to-point (PTP) routing will increase the PRnet throughput, but is highly dependent on the station (at least in the proposed implementation) for route updates. In this appendix we discuss the investigation of strategies for improving conferencing ability without increased (and possibly with reduced) throughput requirements. The emphasis will be on locating any specialized functions away from the station for two reasons:

- (1) The station is already overburdened
- (2) In a tactical network there is the possibility of losing the station (or any site); therefore, these functions must be capable of moving from site to site as required.

A. Unstructured Conferencing

Requiring that a conferencing system be governed by a control protocol can be problematical. For example, it prohibits the use of minimum cost (or miniaturized) single-processor SIUs, such as the standard PRnet terminal interface LSI-11 (TIU), and can lead to unacceptable CONTROL delays if the host network becomes sufficiently unreliable or slow. From a user viewpoint, there is sometimes no need for the control mechanism, as in a small round-table discussion. Besides, complete dependency of the voice transceivers on external control seems unrobust.

A control protocol would not be essential for a working conferencing system if the voice transceivers could operate autonomously (without a VCS or CONTROL* network). To use such a system, each conferee would have to manually dial the others and the VCTs would establish the fully connected VOICE network. Floor handoff would be based on local/foreign speech activity and would be achieved by automatic regulation of the local TRANSMITTER; one simple approach is to activate the TRANSMITTER as long as there is local speech activity, but no foreign speech activity. Of course, any such scheme will have

 * All capitalized words refer to specific program code segments rather than the more general standard meanings.

potential deadlocks that will require that all TALKERS briefly fall silent. A limited feedback facility (see Section C-3) would be helpful and perhaps essential, but smooth TALKER handoff will always require conferee skill and cooperation.

An unstructured conferencing system could be accommodated by a standard PRnet TIU with the addition of a vocoder interface. This could be achieved by using the currently implemented VOICE transceiver with the addition of a very small voice communication overseer (VCO) to the VOICE processor, which would interact with the VCT through the same logical interface used by the VCS. Responding to the user's dialer, the VCO would direct the VCT to establish the VOICE network; using the speech activity signals supplied by the VCT, it would also implement the FLOOR handoff algorithm by enabling and disabling the TRANSMITTER when appropriate.

B. Broadcast Packets

Current conferencing protocols require n times the nominal vocoder bit-rate where n is the number of users. If the PRnet were able to support broadcast packets, with each packet containing a destination list (or each network terminal node bearing a source list), then the bit-rate would not be dependent on the the number of conferees. For example, when only one talker is allowed access at a time, 16-kbits/s CVSD users require 13-kbits/s average (assuming speech activity detection), and 16 kbits/s peak. The current NVCP protocol requires $(n-1)$ times that bit-rate. There have been several studies of broadcast protocols [6]. We propose to review these studies with respect to the speech conferencing requirements.

C. Limited Feedback Conferencing

The speech conferencing protocols currently implemented for the ARPAnet experiments allow only one talker at a time, with first-come-first-served floor handoff. Natural conferencing around a table permits feedback and interruption, and the listeners are capable of hearing both parties simultaneously. Without this feedback, response to other conferees is restricted, making them feel isolated, and resulting in a sequence of weakly connected monologues. Therefore, protocols without an interruption facility have a low user effectiveness.

Limited feedback means that LISTENERS are allowed to make short responses at random times while one TALKER has the floor. Somehow these packet streams must be combined with the TALKER's voice stream. It is suggested that this mixing be done locally at every SIU for best real-time response (it is assumed that the VOICE network is fully connected). Each speech RECEIVER would maintain a small packet queue for all the CONFEREES where response packets could be collected and ordered. Given

appropriate hardware, local users could be presented with an audio mix of the speech streams in real-time (see Appendix D). Lacking such facilities, mixing could be done at the software level: interruption streams would be output to the speech decoder in a round-robin fashion whenever the TALKER falls silent, or for better response, perhaps they could be strategically inserted within the TALKER's voice stream. The second possibility is more natural and would prevent filibustering, but requires some method of determining strategic interruption points. Note that this much of the limited feedback facility is completely within the VOICE protocol and that the hardware and software mixing mechanisms are fully compatible.

To fully implement limited feedback, several extensions to the CONTROL protocol are required. Foremost is the ability to selectively allow LISTENERS to generate up to a specified number of speech packets. A provision for LISTENER requests to interrupt (generated automatically upon local speech activity detection and directed to the CHAIRMAN) would also be desirable, especially for highly structured conferences and to regulate a given LISTENER's frequency of interruption. With these facilities, the CHAIRMAN program could dynamically optimize the amount and source of the interruption traffic in view of the network and SIU resources available for handling the additional load (note that two one-second interruptions occurring every 30 seconds would require only 7% more bit-rate but would dramatically enhance participant interaction).

D. All-to-All Conferencing

In natural face-to-face conferencing, any conferee can talk at will to any other conferee. Teleconferencing in such an all-to-all mode introduces problems, including much higher bandwidth utilization [by a factor of $n(n-1)$, where n is the number of users], and increased hardware cost [$(n-1)$ decoders per site: one for each remote conferee], but significantly increases user effectiveness.

All-to-all conferencing can be accomplished while keeping bandwidth utilization, hardware cost, and complexity within reasonable bounds. The idea is to route all active transmissions to a central conference clearing house, decompress the voice streams, add algebraically, then recompress and transmit (or broadcast) to all CONFEREES. By centralizing this process at strategic points within a network, overall bandwidth utilization would be dramatically reduced, and only one decompressor would be needed at each conference site without sacrificing voice quality.

In the PRnet with hierarchical routing, the station would be one logical place to perform concentration (unfortunately, the station may already be too overburdened with functions to make this practical), while for point-to-point (PTP) routing it would be necessary to situate the concentration at one of the SIUs (requiring that they be functional

repeaters). A more complicated placement would be at a repeater, which requires a change in the hop transport protocol. In other network environments, a GATEWAY would be an ideal concentrator location, especially for internet conferencing applications in which the benefits bestowed by concentration are critical.

The SPS-41 can be used as a time-shared decompression/recompression processor. The current implementation of the CVSD algorithm in the SPS-41 can support a full-duplex 80-kbits/s rate, assuming the simplest implementation wherein all transmissions are concentrated into one compressed message and forwarded to everyone. This processing rate would support 16 10-kbits/s CVSD users, or 10 16-kbits/s CVSD users (assuming equal time for decompression and recompression). Of course, a faster processor such as the FPS could accommodate even more users.

Several problems must be studied before a detailed implementation strategy for concentration can be specified. These include:

- (1) Single-site concentration is not robust, since the site could become unavailable. Consequently, there should be provisions for automatic movement of the logical concentration point when necessary. Unfortunately, such mobility requires a reliable relocation protocol and several concentrators per network.
- (2) There are reordering and mixing difficulties because of real-time asynchrony and random delays in the voice streams; i.e., how do you align the streams to achieve the proper real-time audio synthesis? Appropriate strategies must be developed.
- (3) Centralized concentration adds a store-and-forward delay in addition to SPS-41 processing time, which may be unacceptable for real-time conferencing. The delay could be curtailed by minimizing packet buffering at the concentrators, but this aggravates the reordering/mixing problem.
- (4) Concentration of all packets into one stream will produce an annoying side tone, since each conferee's voice stream is returned to him. To avoid this, the concentrator would have to generate one voice stream for the set of LISTENERS (a mix of all TALKER streams) and a separate stream for each TALKER (one without his own speech); trade-offs incurred include increased bandwidth utilization proportional to the number of TALKERS, and extra processing delay.
- (5) Any concentration mechanism must be compatible with the conference CONTROL and VOICE protocols, so that

conferences can take place in an environment with or without concentration, with the ability to reconfigure as concentration sites change availability status.

Appendix D

INTELLIGIBLE INTERRUPTION TECHNIQUES

Appendix D

INTELLIGIBLE INTERRUPTION TECHNIQUES

In the PRnet, radio receivers are time-shared asynchronously so that the radios can accept multiple signals, limited primarily by contention. This capability makes intelligible interruption conferencing possible if each SIU can combine the VOICE streams as they arrive (see Appendix C). Here we discuss two hardware mixing strategies.

The first approach is to provide a separate decoder for each VOICE stream, summing their analog outputs. This method would provide the best voice quality and natural synthesis of the streams, but incurs two penalties:

- (1) The hardware cost is proportional to the number of conferees. The low cost of CVSD decoders (one chip) makes this a relatively minor point.
- (2) The number of decoders available limits the permitted number of simultaneous responses. A possible solution is to use a software technique (see Appendix C) to map several voice streams onto a single decoder input. In this way, as few as two decoders may suffice: one for the interruption streams, and one for the TALKER's stream (which retains the original voice quality). A bonus is the reduced amount of hardware needed.

The second approach is less costly but sacrifices some voice quality in the interruption channel (the main TALKER channel is the standard CVSD decoder). The interruption channel is generated by bit-wise addition of the binary voice streams at the input of a simple low-pass filter. Since this decoding is linear rather than adaptive, some signal degradation is incurred, but recent experiments at SRI have indicated that intelligibility is surprisingly good. Such a simple summing method is best employed to provide a low-volume interrupt facility in place of flashing lights, as currently available.

REFERENCES

1. D. Cohen, "Specifications for the Network Voice Protocol", NSC Note No. 68, March 1976.
2. J. Forgie, "A Proposal for SATNET Voice Conferencing", NSC Note No. 104, January 1977.
3. J. Sussman, "Point-to-Point Routing Proposal", PRTN Note No. 194, August 1976.
4. D. Cohen, "Internetting, or Beyond NCP", PRTN Note No. 213, February 1977.
5. J. Sussman, "Route Assignment Proposal", PRTN Note No. 192, August 1976.
6. Y. Dalal, "Broadcast Protocols in Packet Switched Computer Networks", Technical Report No. 128, Stanford University, April 1977.
7. R. Sunlin, Private Communication.

DEPARTMENT OF DEFENSE

CONTRACT PRICING PROPOSAL

(RESEARCH AND DEVELOPMENT)

Form Approved

Budget Bureau No. 22-R0100

This form is for use when (i) submission of cost or pricing data (see ASPR 3-807.3) is required and (ii) substitution for the DD Form 633 is authorized by the contracting officer.

PAGE NO.

NO. OF PAGES

NAME OF OFFEROR

SRI INTERNATIONAL

SUPPLIES AND/OR SERVICES TO BE FURNISHED

Research

HOME OFFICE ADDRESS (Include ZIP Code)

333 Ravenswood Avenue
Menlo Park, California 94025

DIVISION(S) AND LOCATION(S) WHERE WORK IS TO BE PERFORMED

As above

TOTAL AMOUNT OF PROPOSAL

See Part Two--Contractual Provisions

GOVT SOLICITATION NO.

- - -

DETAIL DESCRIPTION OF COST ELEMENTS

No. ISU 77-117

1. DIRECT MATERIAL (Itemize on Exhibit A)	EST COST (\$)	TOTAL EST COST ¹	REFER- ² ENCE
a. PURCHASED PARTS			
b. SUBCONTRACTED ITEMS			
c. OTHER - (1) RAW MATERIAL			
(2) YOUR STANDARD COMMERCIAL ITEMS			
(3) INTERDIVISIONAL TRANSFERS (At other than cost)			
TOTAL DIRECT MATERIAL			
2. MATERIAL OVERHEAD ³ (Rate % X \$ base =)			
3. DIRECT LABOR (Specify)	ESTIMATED HOURS	RATE/HOUR	EST COST (\$)
TOTAL DIRECT LABOR			
4. LABOR OVERHEAD (Specify department or cost center) ³	O.H. RATE	X BASE =	EST COST (\$)
TOTAL LABOR OVERHEAD			
5. SPECIAL TESTING (Including field work at Government installations)	EST COST (\$)		
TOTAL SPECIAL TESTING			
6. SPECIAL EQUIPMENT (If direct charge) (Itemize on Exhibit A)			
7. TRAVEL (If direct charge) (Give details on attached Schedule)	EST COST (\$)		
a. TRANSPORTATION			
b. PER DIEM OR SUBSISTENCE			
TOTAL TRAVEL			
8. CONSULTANTS (Identity - purpose - rate)	EST COST (\$)		
TOTAL CONSULTANTS			
9. OTHER DIRECT COSTS (Itemize on Exhibit A)			
10. TOTAL DIRECT COST AND OVERHEAD			
11. GENERAL AND ADMINISTRATIVE EXPENSE (Rate % of cost element Nos.) ³			
12. ROYALTIES ⁴			
13. TOTAL ESTIMATED COST	See Part Two--Contractual Provisions	No. ISU 77-117	
14. FEE OR PROFIT			
15. TOTAL ESTIMATED COST AND FEE OR PROFIT			

This proposal is submitted for use in connection with and in response to (Describe RFP, etc.)

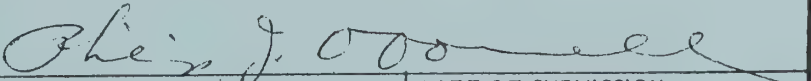
- - -

and reflects our best estimates as of this date, in accordance with the instructions to offerors and the footnotes which follow.

TYPED NAME AND TITLE

Philip J. O'Donnell
Manager
Contract Administration

SIGNATURE



NAME OF FIRM

SRI INTERNATIONAL

DATE OF SUBMISSION

1977 AUG 22

[illegible]

☒ YES ☐ NO *If yes, identify below.*

Defense Contract Audit Agency, SRI Resident

415 326-6200, X2089

☒ YES ☐ NO If yes, identify on a separate page. See Part Two--Contractual Provisions No.

☐ YES ☒ NO if yes, identify: ☐ ADVANCE PAYMENTS ☐ PROGRESS PAYMENTS OR ☐ GUARANTEED LOANS

☒ YES ☐ NO If yes, identify

☒ YES ☐ NO If no, explain on a separate page.

1. The purpose of this form is to provide a standard format by which the offeror submits to the Government a summary of incurred and estimated cost (*and attached supporting information*) suitable for detailed review and analysis. Prior to the award of a contract resulting from this proposal the offeror shall, under the conditions stated in ASPR 3-807.3, be required to submit a Certificate of Current Cost or Pricing Data (see ASPR 3-807.3(e) and 3-807.4).

a. the judgmental factors applied and the mathematical or other methods used in the estimate including those used in projecting from known data, and

b. the contingencies used by offeror in his proposed price.

4. The format for the "Cost Elements" is not intended as rigid requirements. These may be presented in different format with the prior approval of the contracting officer if required for more effective and efficient presentation. In all other respects this form will be completed and submitted without change.

5. By submission of this proposal, offeror, if selected for negotiation, grants to the contracting officer, or his authorized representative, the right to examine, for the purpose of verifying the cost or pricing data submitted, those books, records, documents and other supporting data which will permit adequate evaluation of such cost or pricing data, along with the computations and projections used therein. This right may be exercised in connection with any negotiations prior to contract award.

1 Enter in this column those necessary and reasonable costs which in the judgment of the offeror will properly be incurred in the efficient performance of the contract. When any of the costs in this column have already been incurred (e.g., on a letter contract or change order), describe them on an attached supporting schedule. Identify all sales and transfers between your plants, divisions, or organizations under a common control, which are included at other than the lower of cost to the original transferee or current market price.

2 When space in addition to that available in Exhibit A is required, attach separate pages as necessary and identify in this "Reference" column the attachment in which information supporting the specific cost element may be found. No standard format is prescribed; however, the cost or pricing data must be accurate, complete and current, and the judgment factors used in projecting from the data to the estimates must be stated in sufficient detail to enable the contracting officer to evaluate the proposal. For example, provide the basis used for pricing materials such as by vendor quotations, shop estimates, or invoice prices; the reason for use of overhead rates which depart significantly from experienced rates (reduced volume, a planned major rearrangement, etc.); or justification for an increase in labor rates (anticipated wage and salary increases, etc.). Identify and explain any contingencies which are included in the proposed price, such as anticipated costs of rejects and defective work, or anticipated technical difficulties.

3. Indicate the rates used and provide an appropriate explanation. Where agreement has been reached with Government representatives on the use of forward pricing rates, describe the nature of the agreement. Provide the method of computation and application of your overhead expense, including cost breakdown and showing trends and budgetary data as necessary to provide a basis for evaluation of the reasonableness of proposed rates.

4 If the total royalty cost entered here is in excess of \$250 provide on a separate page (or on DD Form 783, Royalty Report) the following information on each separate item of royalty or license fee: name and address of licensor; date of license agreement; patent numbers, patent application serial numbers, or other basis on which the royalty is payable; brief description; including any part or model numbers of each contract item or component on which the royalty is payable; percentage or dollar rate of royalty per unit; unit price of contract item; number of units; and total dollar amount of royalties. In addition, if specifically requested by the contracting officer, a copy of the current license agreement and identification of applicable claims of specific patents shall be provided.

5 Provide a list of principal items within each category indicating known or anticipated source, quantity, unit price, competition obtained, and basis of establishing source and reasonableness of cost.



STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 · U.S.A.

10 August 1977

Proposal for Research

SRI No. ISU 77-117

PACKET RADIO SPEECH EXPERIMENTAL PROGRAM

Part Two--Contractual Provisions

Prepared for:

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209
Attn: Vinton G. Cerf

SRI Proposal for Research No. ISU 77-117

PACKET RADIO SPEECH EXPERIMENTAL PROGRAM

Part Two--Contractual Provisions

I ESTIMATED TIME AND CHARGES

The estimated time required to complete this project and report its results is 12 months. SRI International could begin work on receipt of a fully executed contract.

Attached are a cost estimate and support schedules in lieu of the DD Form 633-4, pursuant to the provisions of ASPR 16-206.2, and an SRI Form 1642 (based on DD Form 1861), entitled "Contract Facilities Capital and Cost of Money," supporting the amount of inputted interest and capital employed. Also enclosed is a signed DD Form 633-4 complete except to the "Detail Description of Cost Elements."

II GOVERNMENT-FURNISHED EQUIPMENT

SRI International's cost proposal is based on the use of the following equipment that was acquired on previous packet speech ARPA contracts and is expected to be provided as GFE for SRI's use in the packet radio speech experimental program:

- . PDP-11/40 processor with 64K memory, 2 RK05 disks, ARPAnet interface, A/D and D/A, and 2 CVSD hardware interfaces.
- . (2) LSI-11 based dual-processor computer systems with 40K memory, CVSD hardware interfaces, and PRnet interfaces.

III CONTRACT FORM

It is requested that any contract resulting from this proposal be written under cost-plus-fixed fee Basic Agreement No. N00014-76-A-0168 between the Office of Naval Research and SRI International.

If any resulting contract is to be assigned to DCASD for administration, we request that the contract contain provision identifying our address for receipt of remittance as follows:

SRI International
Post Office Box #80163
Worldway Postal Center
Los Angeles, CA 90080

Contracts and other mail should continue to show our mailing address as:

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

IV DATA CERTIFICATION

SRI has not delivered nor is it obligated to deliver to the Government under any other contract or subcontract the same data required under this Request for Quotation.

V ACCEPTANCE PERIOD

This proposal will remain in effect until 1 November 1977. If consideration of the proposal requires a longer period, SRI will be glad to consider a request for an extension of time.

VI SECURITY CLASSIFICATION

SRI International holds a Top Secret facility clearance, which may be verified through the cognizant military security agency, San Francisco Defense Contract Administration Services Region, Attn: Office of Industrial Security, 866 Malcolm Road, Burlingame, California 94010. Staff assignments will be in accordance with the level of security assigned to the work.

VII COST ESTIMATE

The work statement for this proposal contains four tasks. The cost estimates combine the first two tasks and separate the optional tasks three, and four. The costing of the first task is based on performing two CAP tests (see the technical proposal for details) and two mobile speech exercises. Additional CAP tests will cost \$4,337 per test, and additional mobile speech exercises will cost \$5,310 per exercise.

COST ESTIMATE

Task 1 - PRnet Testing and Evaluation Using Voice Traffic
Task 2 - Expanded Internet Services For Voice Traffic

Personnel

Supervisory,	40 hours at \$17.72	\$ 709
Professional,	1478 hours at \$10.63	15,711
Technical,	96 hours at \$ 9.44	906
Clerical,	40 hours at \$ 5.97	<u>239</u>

Total SRI Labor \$ 17,565

Payroll Burden at 31.0% 5,445

Total Labor and Burden \$ 23,010

Research Overhead at 90.0% of Labor and Burden 20,709

G & A at 21.0% of Labor and Burden 4,832

Support Cost

Travel and Subsistence	\$ 2,258
Communications	700
SRI PDP-10 1090T Computer	6,000
Report Preparation	3,158
LSI-11 Maintenance Supplies	<u>700</u>

Total Support Cost \$ 12,816

Support Cost Burden at 3.6% of Support Cost 461

CAS 414 Cost 1,572

Total Estimated Cost 63,400

Fixed Fee 4,755

TOTAL ESTIMATED COST PLUS FIXED FEE \$ 68,155

SRI International
Proposal For Research No. ISU 77-117

COST ESTIMATE

Task 3 - (Optional) Conference CONTROL Protocol Implementation

Personnel

Supervisory, 72 hours at \$17.99	\$ 1,295
Professional, 1,848 hours at \$ 9.12	<u>16,854</u>

Total SRI Labor	\$ 18,149
-----------------	-----------

Payroll Burden at 31.0%	<u>5,626</u>
-------------------------	--------------

Total Labor and Burden	\$ 23,775
------------------------	-----------

<u>Research Overhead</u> at 90.0% of Labor and Burden	21,398
---	--------

<u>G & A</u> at 21.0% of Labor and Burden	4,993
---	-------

Support Cost

Travel and Subsistence	\$ 3,252
PDP-11/40 Maintenance	<u>6,000</u>

Total Support Cost	\$ 9,252
--------------------	----------

<u>Support Cost Burden</u> at 3.6% of Support Cost	333
--	-----

<u>CAS 414 Cost</u>	1,118
---------------------	-------

<u>Total Estimated Cost</u>	60,869
-----------------------------	--------

<u>Fixed Fee</u>	<u>4,565</u>
------------------	--------------

TOTAL ESTIMATED COST PLUS FIXED FEE	\$ 65,434
-------------------------------------	-----------

COST ESTIMATE

Task 3a - (Optional) LSI-11 Speech Terminal Purchase

Personnel

Technical,	80 hours at \$ 8.28	\$	<u>662</u>
Total SRI Labor		\$	662
Payroll Burden at 31.0%			<u>205</u>
Total Labor and Burden		\$	867
<u>Research Overhead</u> at 90.0% of Labor and Burden			780
<u>G & A</u> at 21.0% of Labor and Burden			182

Support Cost

Dual-Processor LSI-11 Purchase	\$	8,631
Materials and Supplies (see Schedule D)		200
		<u> </u>
Total Support Cost	\$	8,831
<u>Support Cost Burden</u> at 3.6% of Support Cost		318
<u>CAS 414 Cost</u>		47
<u>Total Estimated Cost</u>		11,025
<u>Fixed Fee</u>		<u>551</u>
TOTAL ESTIMATED COST PLUS FIXED FEE	\$	11,576

SRI International
Proposal For Research No. ISU 77-117

COST ESTIMATE

Task 4 - (Optional) CVSD Hardware Construction

Personnel

Professional,	693 hours at \$10.31	\$	7,145
Technical,	154 hours at \$ 8.33	\$	<u>1,283</u>

Total SRI Labor	\$	8,428
-----------------	----	-------

Payroll Burden at 31.0%	<u>2,613</u>
-------------------------	--------------

Total Labor and Burden	\$	11,041
------------------------	----	--------

<u>Research Overhead</u> at 90.0% of Labor and Burden	9,937
---	-------

<u>G & A</u> at 21.0% of Labor and Burden	2,319
---	-------

Support Cost

Material and Supplies (see Schedule D)	\$	8,064
---	----	-------

Total Support Cost	\$	8,064
--------------------	----	-------

<u>Support Cost Burden</u> at 3.6% of Support Cost	290
--	-----

<u>CAS 414 Cost</u>	521
---------------------	-----

<u>Total Estimated Cost</u>	32,172
-----------------------------	--------

<u>Fixed Fee</u>	<u>2,091</u>
------------------	--------------

TOTAL ESTIMATED COST PLUS FIXED FEE	\$	34,263
-------------------------------------	----	--------

SCHEDULE A - DIRECT LABOR

Direct labor charges are based on the actual salaries for the staff members contemplated for the project work plus a judgmental factor of base salary for merit increases during the contract period of performance. The precise factor applied is dependent on the estimated period of performance. Frequency of salary reviews and level of merit increases are in accordance with SRI's Salary and Wage Payment Policy as published in Topic No. 505 of the SRI Administration Manual and as approved by the Defense Contract Administration Services Region.

SCHEDULE B - INDIRECT BIDDING RATES

These rates have been accepted by DCASMA, San Francisco, as interim 1977 bidding and billing rates. We request that these rates not be specifically included in the contract, but rather that the contract provide for reimbursement at billing rates acceptable to the Contracting Officer, subject to retroactive adjustment to fixed rates negotiated on the basis of historical cost data. Included in payroll burden are such costs as vacation, holiday and sick leave pay, social security taxes, and contributions to employee benefit plans.

SCHEDULE C - MATERIALS AND SERVICES

Travel

TASK 1 and 2:

3 trips to Washington, D.C. at \$392	\$ 1,176
--------------------------------------	----------

4 trips to Los Angeles, Calif. at \$51	204
--	-----

Subsistence:

9 days at Washington, D.C. at \$42.50/day	383
---	-----

8 days at Los Angeles, Calif. at \$30/day	240
---	-----

Car Rental:

17 days at \$15.00/day	255
------------------------	-----

TASK 3:

6 trips to Washington, D.C. at \$392	\$ 2,352
--------------------------------------	----------

Subsistence:

18 days at Washington, D.C. at \$42.50/day	765
--	-----

Car Rental:

9 days at \$15.00/day	135
-----------------------	-----

Air fares and car rental rates are based on prices established in current Official Airline Guide. Domestic subsistence rates and travel by private auto are quoted at published standards.

Computer Time

Computation for the proposed project will be done on a DEC system 1090T computer facility located at SRI. The 1090T is a large timesharing system that can support upwards of 100 simultaneous users who are mainly doing interactive program development and debugging. The system is connected to other computers throughout the country via the ARPANET as well as ordinary telephone lines.

For interactive use, SRI projects subscribe for access to a share of the facility, measured in Computer Resource Units (CRUs) over a given period of time. A CRU is defined to always deliver at least 1/2% of a 1090T CPU.

For SRI projects sponsored by the U.S. Government, the charge for computer service is \$500. per month per CRU. This charge is based upon the allowable costs of the computer facility, and will be adjusted periodically to reflect actual cost experience. We estimate that the project proposed here requires a total cost of 1090T computer time of \$6,000 for 1 CRU over the 12 month contract period.

Communications

This is an estimate of the toll charges for telephone calls during the period of performance.

SCHEDULE D - EQUIPMENT, MATERIALS AND MAINTENANCE

The following equipment costs are based upon formal quotations from previous suppliers and engineering estimates based on the most current price sheets.

TASK 1:

- . Maintenance materials for two existing LSI-11 systems \$ 700

TASK 3:

- . PDP 11/40 Maintenance 8,000

TASK 3a:

- . LSI-11 based dual-processor packet speech interface system with 40K memory, inter-processor communication, CVSD interface, PRnet interface, and terminal/loading interfaces-complete system quote. 8,220
- . Twelve month extended warrantee for the above system. 411
- . Material for patch panel construction 200

TASK 4:

- . Materials for construction of CVSD hardware. 8,064
 - Four units, each unit will require:
 - MLSI-1710 foundation module \$ 165
 - DRV11-B DMA interfaces (2) 1,127
 - Integrated circuits 498
 - PROM memory 198
 - Misc. (connectors, cables, etc.) 28

SCHEDULE E - REPORT COSTS

Report costs are estimated on the basis of the number of pages of text and illustrations and the number of copies of reports required in accordance with the following rates per page.

Editing	\$ 4.12
Composition	3.83
Coordination	1.12
Proofreading	1.68
Illustrations:	
Simple	13.21
Average	20.92
Complex	40.74
Press and Bindery	0.033 per impression

Following is a cost breakdown of the estimated cost of report production:

Printing, 210 pages at \$10.75 per page (including editing, composition, report coordination, and proofreading)	\$ 2,258
Illustration, 30 pages at \$20.92 per illustration	\$ 628
Press, bindery, and photography for 8,250 printed pages at \$0.033 per printed page	\$ 272

CONTRACT FACILITIES CAPITAL AND COST OF MONEY

CLIENT ARPA		PROPOSAL NO. ISU 77-117, Task 1 & Task 2	
1. OVERHEAD POOLS		2. COST ESTIMATE REFERENCE	3. CONTRACT OVERHEAD ALLOCATION BASE
		FACILITIES CAPITAL COST OF MONEY	
		4. FACTORS	5. AMOUNT
Research Overhead (TL & B)		\$23,058	.04384 1011
G & A (TL & B)		23,058	.00284 65
Computer Center (Computer \$)		6,000	.08209 493
Support Cost Burden (OSC \$)		6,816	.00080 6
6. CONTRACT FACILITIES CAPITAL COST OF MONEY		\$1,575	
7. FACILITIES CAPITAL COST OF MONEY RATE		÷ .0775	
8. CONTRACT FACILITIES CAPITAL EMPLOYED		\$20,323	

CONTRACT FACILITIES CAPITAL AND COST OF MONEY

CLIENT		PROPOSAL NO.		
ARPA		ISU 77-117 Task 3		
1.	OVERHEAD POOLS	2.	3. CONTRACT OVERHEAD ALLOCATION BASE	FACILITIES CAPITAL COST OF MONEY
			4. FACTORS	5. AMOUNT
	Research Overhead (TL & B)		\$23,775	.04384 1042
	G & A (TL & B)		23,775	.00284 68
	Computer Center (Computer \$)		-	.08209 -
	Support Cost Burden (OSC \$)		9,252	.00080 8
6. CONTRACT FACILITIES CAPITAL COST OF MONEY			\$ 1,118	
7. FACILITIES CAPITAL COST OF MONEY RATE			÷ .0775	
8. CONTRACT FACILITIES CAPITAL EMPLOYED			\$14,426	

CONTRACT FACILITIES CAPITAL AND COST OF MONEY

CLIENT		PROPOSAL NO.		
ARPA		ISU 77-117 Task 3A		
1. OVERHEAD POOLS	2. COST ESTIMATE REFERENCE	3. CONTRACT OVERHEAD ALLOCATION BASE	FACILITIES CAPITAL COST OF MONEY	
			4. FACTORS	5. AMOUNT
Research Overhead (TL & B)		\$ 867	.04384	38
G & A (TL & B)		867	.00284	2
Computer Center (Computer \$)		-	.08209	-
Support Cost Burden (OSC \$)		8,831	.00080	7
6. CONTRACT FACILITIES CAPITAL COST OF MONEY			\$ 47	
7. FACILITIES CAPITAL COST OF MONEY RATE			÷ .0775	
8. CONTRACT FACILITIES CAPITAL EMPLOYED			\$606	



STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 · U.S.A.

31 December 1975

Proposal for Research

SRI No. ISU 75-184

EVALUATION OF SPEECH TRANSMISSION
OVER PACKET SYSTEMS

Part One--Technical Proposal

Prepared for:

Information Processing Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attention: Mr. Robert Kahn

Prepared by:

Richard W. Becker
Senior Research Psychologist
Sensory Sciences Research Center

Approved:

Karl D. Kryter, Director
Sensory Sciences Research Center

Bonnar Cox, Executive Director
Information Science and Engineering Division

015

APPA Cont No.

100 E

100 E

Proposal for Research No. ISU 75-184

EVALUATION OF SPEECH TRANSMISSION
OVER PACKET SYSTEMS

Part One--Technical Proposal

I INTRODUCTION

This proposal describes a program for evaluating speech communication over digital packet communication networks. A series of experiments is described in which personnel of the Sensory Sciences Research Center (SSRC) of Stanford Research Institute (SRI) will measure the effects of temporal distortions on the quality of speech communication. Specifically, we will measure the effects of introducing interruptions or gaps into speech communication, the effects of losing segments of the speech, and the effects on listeners of the speech being played with segments interchanged. We will measure these effects on high-quality speech and on speech that has been processed by bandwidth-reduction systems similar to those expected to be used in future DoD communication systems. These measurements will be made both by use of a packet radio network (PR net) test bed in Menlo Park, California and by simulation of packet communication system effects on our computer facilities. We will also assess the effects of various gap-filling strategies on the quality of speech. Finally, we will determine the effects of the speech degradation on ability to perform specific two-way communication tasks accurately and efficiently.

This proposal consists of eight sections. Section II, Background, explains the advantages of transmitting speech via digital rather than analog means and the particular advantages of using packet networks for the transmission of digitally formatted speech. Section II also discusses the problems that may occur in using such networks and the research on these problems that has been conducted to date. Section III describes the objectives of the research and evaluation program proposed here. Section IV, Method of Approach, specifically describes the details of six proposed experiments and explains the choice of various parameters and measurements. Section V is a proposed Statement of Work for this program. (Detailed estimates of time and costs to conduct the program are contained in Part Two, Contractual Provisions.) A Management Plan and Description of Facilities make up Sections VI and VII.

This research program is important for the following reasons: First, the potential advantages of digital speech transmission over packet networks make this switching the method of choice for communication networks of the future, particularly for military communications. Second, very little research has been done to determine the effects of temporal distortion of speech--for example, the introduction of gaps into speech--on the quality of speech communication. No research has been done on the effects of the temporal distortion to be expected in packet networks. Nor has any research been done on the effects of the combined temporal and bandwidth distortion that can be expected in future communication networks. Finally, no research has been done on the effects of this degradation on the accuracy and efficiency of two-way transmission of information.

By conducting this program of research we will extend scientific knowledge of the effects of the interaction of various temporal-domain and frequency-domain distortions of speech. This knowledge will enable us to predict, for given configurations of networks and of transmission protocols within a network, the effects of packet networks on the quality of the speech transmitted over them and on the ability of humans to use them. Thus timely execution of this research will enable choices in network construction to be made that will result in minimal degradation of human communication performance.

The Sensory Sciences Research Center (SSRC) of SRI is uniquely qualified to conduct this evaluation program, for several reasons. First, the SSRC has a long history of experience in the assessment of the intelligibility and quality of communication systems [Kryter et al., 1962; Clarke et al., 1965; and Becker and Clarke, 1965],* and has particular experience [Becker and Kryter, 1975] in the evaluation of low-bandwidth digital systems developed under the ARPA program. Second, the management and integration of the packet radio test-bed project in Menlo Park is being performed by SRI. Thus, the proposed program of research will be part of an integrated plan comprising: (1) implementation of a PR net test bed, by the Telecommunications Sciences Center (TSC) of SRI; (2) implementation of speech transmission on this network by the TSC; (3) measurement of the quality of speech transmission over this network, through the combined efforts of the TSC and the SSRC; and (4) prediction by the SSRC, of the quality of speech communication and performance under a variety of future possible PR net configurations.

*References are listed in Section VIII.

II BACKGROUND

A. General

There are many reasons for wanting to transmit speech in digital rather than analog format. Not the least of these is the ability to repeatedly regenerate the signal without accumulating distortion in the transmission process. Problems in switching and timesharing are also amenable to straightforward solution in the digital format. Perhaps of prime importance in DoD applications is the much greater security obtainable by digital transmission.

Digitally encoded information can be transmitted between two sites in various ways. Of particular interest is the use of packet networks for the transmission of such information. By breaking each message up into small packets, such networks have great potential for improving the use of limited channel resources--such as RF bands or telephone lines--by sharing these resources among many users rather than dedicating them to particular users. Since ARPA is now supporting the development of packet digital communication networks for nonspeech data transmission, it would be desirable if these networks could also be used for speech transmission.

However, the use of digital format for speech transmission introduces several problems. A major difficulty in using digital speech is the high bandwidth required for transmission. Even telephone-quality speech (band-limited to 3000 Hz) when sampled at the Nyquist rate (6000 samples/sec) and quantized to 6 or 7 bits may require--without further coding--a bandwidth on the order of 50,000 Hz [Flanagan, 1972]. If one requires very high-quality speech (for example, 6000-Hz bandwidth, 8-bit quantization), the required transmission bandwidth may be doubled. To conserve bandwidth, it thus becomes necessary to process speech via bandwidth-reduction techniques. Such techniques often degrade the intelligibility and quality of speech.

When the digitized speech is transmitted over a packet network, further problems are introduced. Packet switching introduces inherent delays in the communication process and also introduces possibilities for error, such as packets arriving out of sequence, and packets needing to be retransmitted. Strategies that are intended to correct these errors, used at forwarding points and at the receiver terminal, typically introduce further delay. This time-domain distortion may greatly degrade the quality of spoken communication. It is therefore very important to evaluate how well speech can be transmitted over such networks under various test-bed configurations and various communication task constraints.

The remaining parts of this section describe in further detail the problems that may arise because of temporal-domain and frequency-domain distortion in packet digital speech transmission, and the research which has been conducted on these problems.

B. The Temporal-Distortion Problem

A packet communication network breaks up a message into small packets (typically on the order of 1000 bits) and sends the packets from a source to a destination, typically through a series of repeater stations, to be reassembled into the message by the destination station. Error-free transmission can be accomplished through error-checking coding and retransmission of packets that are found to have errors. Obviously, such a process may introduce delays into the communication process. These delays occur as the result of the initial collection of a packet before transmission; the time required to retransmit some packets; and, when the system is heavily loaded, the unavailability of routes.

The effects of such delays on speech communication depend on the length and distribution of the delays, the reassembly strategies used at the destination station, and the kind of communication task.

In the evaluation of various transmission and reconstruction strategies, the major trade-off is between what we shall call communication lag and interruption rate. Communication lag is the length of time between when a message is originated by the speaker and when it is first heard by the listener. Two parameters are associated with this variable. One is the duration between the beginning of the spoken message and the beginning of the heard message; the other is the duration between the end of the spoken message and the end of the heard message. The values of these parameters will depend on the transmission and reconstruction strategies used and on the configuration of the network during transmission. In general, the lags will not be constant but rather will be described by a probability distribution function with an expected value, variance, skewness, and so forth. In one extreme example of transmission and reconstruction strategies, a packet of 1000 bits would simply be sent as soon as it had been collected at the transmission terminal and be put out to the listener as soon as it had been collected at the receiving terminal. In this case, the mean communication lag between the beginnings of the transmitted and the received messages would be the mean time of transmission through the network under its current configuration and load. Under these circumstances, the expected value of the lag between end of spoken message and end of heard message would be the same.

If, however, we modify this transmission strategy so that a packet is not transmitted until verification has been received that the previous packet has been successfully transmitted and played out to the listener, then the mean lag between beginning of spoken message and beginning of heard message remains the mean time for transmission through the network. However, the expected time between end of spoken message and end of heard message could be much longer, depending on the time required for the successful transmission of each packet and verification of the transmission.

At the other extreme of transmission-reconstruction strategies would be the case in which a message would be collected in its entirety at the receiving terminal (its end being indicated by an end-of-message packet) before being played to the listener. In this extreme case, both the beginning-of-message lag and the end-of-message lag would be equal to the duration of the message plus the mean duration for transmission of a packet through the given network.

The effect of communication lag on communication performance (by which we mean the efficiency, accuracy, and ease of usage) by humans depends on the nature of the communication task. At one extreme, a strictly one-way communication, such as leaving a message for someone to listen to later, the effect would be essentially zero. At the other extreme, two-way interchange of information requiring much feedback and verification in a noisy environment, the effect would be considerable. We would expect the time needed to achieve an accurate interchange of information to be much greater than that for a standard telephone conversation, simply because of the requirement that a message not be started at the listener end until its completion at the speaker end. Furthermore, we would expect this requirement to impose a considerable burden on those talking, forcing them into a very unnatural mode of communication, with resulting loss of efficiency.

A second variable that affects communication performance is interruption rate. This variable is described by the durations of uninterrupted speech, together with the durations of the intervals between such speech. If, for example, a transmission-reconstruction strategy were used in which a packet was not sent until transmission of the previous packet had been finished, the listener would hear: a packet of speech, silence for an interval while the verification was being transmitted and the next packet was being transmitted, the next packet of speech, an interval of silence, and so on. Depending on the relative durations of the uninterrupted speech and of the intervals of silence, communication performance might be affected very little or be totally destroyed.

Because temporal distortion of this kind has not been characteristic of analog transmission systems, or even of circuit-switched digital communication systems, virtually no research has been conducted on the effects of such distortion on the intelligibility and quality of speech communications. However, certain related research suggests that the effects of this distortion are potentially very large.

For example, Miller and Licklider [1950] conducted experiments in which segments of an utterance were removed and replaced by silent intervals. They found that, depending on the percentage of speech they discarded and the frequency with which this was done, single-word intelligibility in high-quality speech was decreased as much as 50%. Fairbanks, Guttman, and Miron [1957] found that, if small portions of the utterance were discarded and the remaining parts were joined rather than being separated by gaps, thereby increasing the speed of presentation of the utterance, comprehension of spoken passages was decreased. Discarding 60% of the passage resulted in a 50% decrease in comprehension. Other investigators [Schubert and Parker, 1955; Huggins, 1972] have found that, if a spoken passage is alternated from ear to ear, some rates of alternation greatly reduce the intelligibility of the utterance.

All this research shares the characteristic that the speech to a single ear did not contain the entire utterance; that is, parts of the utterance were actually missing. In addition, the Fairbanks study not only discarded part of the speech but also presented it at a more rapid rate than that at which it had been uttered, which would be expected to reduce comprehension. These studies are not representative of the conditions that are expected to occur in packet network communication.

The temporal distortion in packet networks will typically consist of presenting the entire utterance to one or both ears, while only introducing gaps in the message rather than actually discarding parts of it.

It is reasonable to expect that, if we actually discard parts of a spoken utterance, single-word intelligibility will be degraded, since the discarded parts will potentially include whole phonemes or parts of words. One might not expect this effect if the entire utterance were presented even though interrupted by periods of silence. The only published reference to such a study is by Huggins [1972] in which he reported that intelligibility of connected speech can be destroyed by introducing 200 msec gaps every 20 msec of speech or by introducing 120-msec gaps every 62 msec of speech. These conclusions were verified by tests at SRI under a current research project [Becker and Kryter, 1976].

It should also be noted that even these experiments, which did not discard segments of speech but rather introduced gaps in the speech, were all done with a fixed amount of interrupted speech, a fixed interruption, and a fixed frequency of interruption. Thus none of these experiments represented the situation in which the mean durations of speech, mean durations of gaps, and mean frequency of interruption are specified but the actual values are varied according to some probability distribution.

In summary, evidence suggests that interruptions of speech may have disastrous effects on the intelligibility and quality of speech communication. Unfortunately, the values of interruption rate and interruption amount that have been shown to have significant effects on speech communication lie well within the boundaries of values that might be expected in PR nets.

Therefore, it is of critical importance to formally determine exactly how intelligibility and quality vary as a function of these parameters, in order to recommend network strategies that will minimize undesirable effects. It is equally important to determine how these temporal distortion effects affect speech that is already degraded by bandwidth-reduction processes, bad speech-to-noise ratios, low bandwidth, and so forth. No research in this area has been conducted, and such research is very important for determining strategies in the construction of future packet communication networks.

C. The Bandwidth Problem

Because of its enormous bandwidth requirement, straightforward pulse-code-modulation (PCM) digital speech has rarely been used in practical situations. Instead, this bandwidth requirement has led to a major effort, beginning with work of Dudley [1939], to process speech in some way that would require less bandwidth. It has long been known that speech is very redundant at both the linguistic level and the acoustic microstructure level. Thus the value of a sample of speech at one moment in time is not independent of its values at surrounding times. Furthermore, a description of one pitch period of a steady-state vowel is a very good descriptor of subsequent pitch periods of the vowel, and so on.

These facts have naturally led to two different kinds of attempts to reduce bandwidth. One set of methods relies on the lowest-level microstructure redundancy and simply recodes the speech in a way that minimizes this redundancy. Examples of such methods are delta modulation,

adaptive delta modulation (ADM), and continuous-variable-slope-delta modulation (CVSD). The other set of methods attempts to use knowledge about the production and perception of speech so as to model the signal and then to extract and transmit only the characteristics needed for high-quality speech. Processes in this set include the formant vocoder, the pattern vocoder, and the linear-predictive-coefficient (LPC) processor. In some cases, the redundancy-reducing coding methods have been applied to the model processors.

Although these processing methods have generally been capable of transmitting understandable speech at reasonably low bit rates (1600-16,000 bits per second), this has been accomplished at the cost of lower intelligibility, unnatural-sounding speech, occasional loss of speaker identity, and lack of resistance to degradation in presence of noise.

For example, in our current evaluation program we have found considerable differences in the various systems' resistance to noise.

Table 1 shows that the LPC processors developed under ARPA sponsorship go from an intelligibility of about 87% at a +26-dB speech-to-noise ratio (SNR) to about 56% at a +3-dB SNR. This is to be compared with a variation from 96% to 76% for a high-quality analog reference system at the same SNRs and to 82% and 70% intelligibility at these ratios for a CVSD system operating at 16,000 bits per second.

Table 1 also shows talker recognition scores for representative bandwidth-reduction systems. Talker recognition or talker discrimination tests are useful for measuring the naturalness of transmitted speech. Although the results of these tests are correlated with intelligibility scores, some systems have very poor talker recognition quality while retaining reasonable intelligibility. Table 1 indicates that all the tested systems showed degradation in this quality, with only the speech of the LPC system being at all comparable to unprocessed speech.

In summary, the bandwidth reduction processing that probably must be used in current digital packet systems significantly degrades the quality of speech. Furthermore, in Section B we have seen that temporal distortions may significantly degrade the quality of high-bandwidth speech. It therefore becomes important to determine exactly what are the effects of each of these distortions and how they interact.

Table 1

PERCENTAGE INTELLIGIBILITY AND TALKER RECOGNITION
OF BANDWIDTH REDUCTION SYSTEMS

System	Percent Intelligibility			Talker Recognition with SNR=26
	SNR=26	SNR=8	SNR=3	
Analog reference	96%	84%	76%	89%
CVSD 16 kbit	82	74	70	--
CVSD 9.6 kbit	72	72	68	72
LPC 3.5 kbit (CH)	87	63	56	83
LPC 3.5 kbit (LL)	86	63	54	78
EXP-Ch.Voc. 4.8 kbit	86	65	51	60
HY2-Ch.Voc. 2.4 kbit	81	57	47	58

D. The Measurement of Quality

The overall merit of a communication system must ultimately be measured by the degree to which the system transmits the same kinds and quantities of information as are communicated in an ideal communication system such as face-to-face conversation in a quiet environment.

Some research has been conducted with the objective of measuring the quality of a communication system by using a single overall index of merit. Preference scaling is the major example of this approach [Munson and Karlin, 1962; Hecker and Williams, 1966]. Presumably, such unidimensional scales have proved of value in the comparative evaluation of systems, but they provide little information about the specific properties and capabilities of the systems so evaluated.

An alternative approach is to consider separately the different kinds of information that can be transmitted by speech. It seems unquestionable that the most significant information that must be transmitted is the strict semantic content of a message. The communication of this information has usually been tested by the use of single-word intelligibility tests. Tests that have been developed include Phonetically Balanced Word Lists (PBs) [Egan, 1948]; the Modified Rhyme Test (MRT) [House et al., 1965]; the Phonetically Balanced Rhyme Test (PBRT) [Clarke and Becker, 1969]; and the Diagnostic Rhyme Test (DRT) [Voiers, Sharpley, and Hehmsoth, 1973]. Each of these tests has advantages and disadvantages. For reasons of ease of administration, overall standardization, and widespread usage, the MRT has been used in our current research and will be used in the research described in this proposal.

However, in addition to intelligibility, investigators have examined many other kinds of information that can be transmitted by a person's speech. A few of the attributes that have been investigated include: the identity of the talker [Becker and Clarke, 1965; Pollack, Pickett, and Sumbly, 1954] the age of the talker [Ptacek and Sander, 1966]; the sex of the talker [Ingemann, 1968] the personality of the talker [Markel, Eisler, and Reese 1967] and the emotional state of the talker [Fairbanks and Hoaglin, 1941]. Although there is some evidence that listeners can do better than chance in correctly identifying some of these traits from spoken utterances alone, performance tends to be rather poor. Nevertheless, some attributes have been identified with sufficient accuracy to warrant development of tests to measure how well a particular communication system transmits the properties of speech necessary to identify those attributes. Examples of attributes that have been incorporated in tests include: talker recognition and discrimination [Becker and Clarke, 1967]; Semantic Differential ratings of voices [Becker and Clarke, 1967; Voiers, 1964]; Psychoacoustic Ratings

of voices [Becker and Clarke, 1967; Holmgren, 1963]; and emotional state as portrayed by actors [Becker and Clarke, 1967]. It has been found that, for analog transmission systems, the ability to transmit one kind of information has been highly correlated with the ability to transmit the other kinds of information. That is, if System A had a better ability to transmit single words intelligibly, it also had a greater ability to transmit the talker's identity, and so on. The advent of digital communication has changed this situation somewhat. Becker and Clarke [1967] reported a case in which a low-bandwidth system scored only 2% better on intelligibility than a vocoder system while scoring 35% better on talker recognition. In our current program of research, similar results have been obtained. Table 1 shows that, although the CVSD 9.6-kbit system is markedly poorer in intelligibility than the channel vocoders, it is definitely superior in transmission of the information necessary to identify talkers.

These results suggest that, in evaluating the quality of speech transmitted over digital communication systems, one should measure more attributes of the speech than single-word intelligibility. Because talker discrimination has proved to be independent of intelligibility in the past, and because such tests are relatively easily administered and can be objectively scored, we will use such tests in the research described in this proposal.

There are many ways of testing talker recognition or discrimination. In the most straightforward method, the one used in obtaining the results given in Table 1, the voices of several people (16 in the case of Table 1) are processed and then played to listeners who know the talkers, to determine whether these listeners can correctly identify the talkers. Although this method has great face validity, it suffers administratively in that it is often difficult to find a suitable group--of talkers and listeners who know them--who are willing to participate in such a test.

An alternative to this method is a talker discrimination test. One such test is a same-different test in which the voices of talkers are processed by a system, and listeners are then presented with pairs of voices that are either from the same talker or different talkers. The task of the listeners is to state whether the voices represent the same talker or different talkers. Because such a test can be administered to listeners who are not familiar with the talkers, it has much to recommend it. The test can be used to determine whether a given system transmits information that can be used to discriminate among talkers, but it has some disadvantages. Possibly, a communication system can transmit the information necessary for discrimination among talkers but also systematically change the attributes of talkers. That is, even though the

talkers' voices may be so changed that people familiar with their natural voices would not recognize them, the systematic transformation of the voices may result in high discrimination scores on this same-different test. For this reason, another kind of test has been constructed. The ABX test presents listeners with three voices: the voice of Talker A, the voice of Talker B, and the voice of Talker X, who is either A or B. The task of the listener is to state which. If this test is administered by processing all three voices through a system to be tested, it suffers the same disadvantage as the same-different test. However, the test can also be administered in the form where the voices of Talker A and Talker B are not processed but the voice of X is. That is, the listener hears the natural voices of A and B, and then hears the voice of one of them after processing by the system being tested. The scores in such a test are much more indicative of the degree to which users of the system will be able to recognize talkers known to them, and also of the degree to which the system can be expected to sound natural to the users. For these reasons, this is the test format that will be used in this research program.

Another method of assessing the quality of a communication system is to determine the accuracy or efficiency with which some specified task requiring two-way communication can be accomplished. Several such tasks have been proposed. For example, Richards and Swaffield [1959] compared the time required by a listener to identify a figure in their conversation test with the time required by the same listener to identify a similar figure over a high-quality system. They found that this ratio was highly correlated with the estimated effort required to understand sentences, which, in turn, was correlated with single word intelligibility. However, the only kind of degradation that was investigated was the masking of speech by various degrees of white noise. It is reasonable to believe that the distortions introduced by digital communication systems, particularly the temporal distortions introduced by packet systems, may degrade communication performance in ways that are not correlated with single-word intelligibility.

Although several different tasks have been suggested for such testing, no standardization has occurred in this area. In our current research program we have constructed two psychoacoustic consoles that can be used for a variety of communication tasks. These consoles and tasks were adapted from the work of Cohen et al. [1973], who used the consoles to measure the effects of environmental noise on task performance. The consoles (represented schematically in Figure 1) consist of a vertical array of four lights and a horizontal array of four digits.

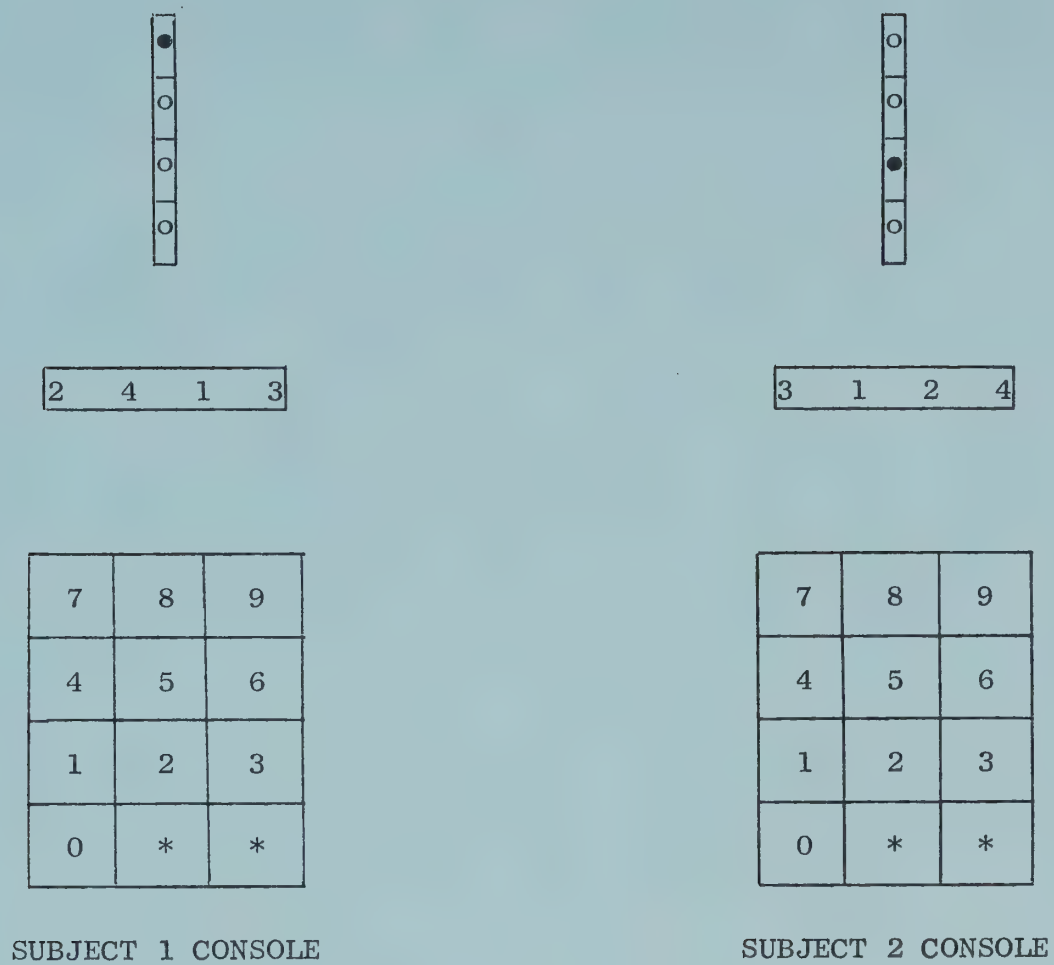


Figure 1 Schematic representation of consoles
used in information exchange tasks

The displays on these consoles are controlled by a PDP-11/10 computer, which can also read the response of each subject and measure the time required to make the response. The tasks that can be performed on these consoles vary from quite simple to very complex. A typical experiment will require Subject 1 to inform Subject 2 which of his vertical lights is lit. Subject 2 will then tell Subject 1 what digit (1 through 4) corresponds to that position on his console. Subject 2 will then tell Subject 1 which light is lit on his vertical display. Subject 1 will then tell Subject 2 what digit is in that position on his digital display. At this point, both subjects key in the sum of the two digits. The displays are then changed: New vertical lights are lit, and new sets of horizontal digits are displayed.

The main virtue of these tasks is that the amount of information that must be exchanged to complete them successfully can be precisely controlled. It is thus possible to systematically investigate communication performance as a function of both channel degradation and complexity of task, as measured by the amount of information that must be exchanged.

In summary, we have indicated that there are many possible ways to measure the quality of a communication network. We have described three particular methods, single-word intelligibility, talker discrimination, and two-way information exchange, and have explained why we choose to use these methods in the experiments described in Section IV.

III OBJECTIVE

The objective of the proposed research is to systematically investigate the effects of certain time-domain distortions--the introduction of gaps and the loss of speech segments--on the perception of speech. Measurements of single-word intelligibility will be made through MRT tests, of talker discrimination through ABX tests, and of efficiency of information exchange through two-way task performance tests. The results of this research should enable prediction of the degradation of human communication performance that will be encountered in a communication network with known bandwidth reduction and temporal distortion parameters.

IV METHOD OF APPROACH

This section describes a series of experiments intended to systematically investigate:

- The effects of time-domain distortion on the quality of high-bandwidth speech.
- The effects of time-domain distortion on the quality of speech that has been processed by bandwidth-reduction systems similar to those to be used in future DoD communication networks.

These effects on quality will be measured by:

- Single-word intelligibility scores obtained through the use of the MRT.
- Talker discrimination scores measured by ABX tests.
- Efficiency of two-way interchange of information as a function of task complexity in controlled information content tasks.

The kinds of temporal distortion that will be introduced will be representative of the temporal distortion to be expected in the use of a hierarchical packet radio communications network for speech transmission coexistent with other data transmission. The properties of this network are:

- Speech is digitized and divided into packets 128 to 2000 bits long. Each packet is sent to a repeater station over an RF link with 100-kbit bandwidth.
- The repeater checks the packet for errors. If the packet is error free, it is sent on to another repeater or the central station over an RF link with 400-kbit bandwidth.
- The packet ultimately reaches the central station (which controls the flow of the packet through the network) and is sent to the receiver terminal, possibly through a series of repeaters, with error checking occurring at each hop.
- When the packet has been correctly received, an acknowledgment is sent back to the originating terminal.

- At each stage in its travel, if a packet is not received correctly, because of either channel error or network contention, the packet must be retransmitted.

The effects of these properties on speech will be the following. All packets will be delayed in reaching the listener by at least the time required to collect a packet of speech plus the minimum time for transmission through the network. The transmission time will not be significantly affected by the propagation time through the physical media, but rather will be a function of the time required for processing each packet at each repeater and the requirement for retransmissions in case of channel errors or interference between two or more messages at a repeater. Because the transmission time will vary from packet to packet, the received speech may be affected in different ways, depending on the strategies used at the sending and listening terminals:

- There may be interruptions in the received speech when a packet arrives significantly later than the previous packet.
- Speech segments may be played out of order when packets do not arrive in sequence.
- Some speech segments may never be played out if they arrive out of sequence.
- To reduce some of these effects, a very large fixed delay may have to be introduced between the utterance of a message and the hearing of it by a listener.

Tests to determine the effects of these degradations on the quality of speech could potentially be done on the actual packet radio test bed. However, there are both theoretical and practical reasons for simulating the network effects for these tests. First, such tests will be time-consuming and would require dedicated use of the packet test bed; such dedicated usage could interfere with other planned development and tests on the test bed. Second, it is very desirable to have complete knowledge of the conditions under which the network is operating during any given test or test sequence. Such knowledge or a guarantee of unchanging conditions cannot be ensured under "live" usage; for example, multipath interference could change considerably during the operation of a test.

For these reasons, most communication tests will be conducted by using simulated network conditions. It should be noted that we are not proposing to simulate the network itself. Rather we plan to use a PDP-11/40-SPS-41 system to simulate the effects of such a network. For

one-way communication tests, the system will be used to perform the required bandwidth-reduction processing of stored speech and to introduce time-domain distortion representative of that which would occur in a packet radio network. In the two-way communication tests, the system will be modified to provide two complete full-duplex A/D-D/A channels with CVSD speech processors at both ends of the communication link. To ensure that these experiments are meaningful, it will be necessary to obtain valid statistics and measurements on distributions of delay times, lengths of uninterrupted messages, and other data. Further, it is desirable to have these statistics for many different configurations of the packet radio test bed. For instance, we would like to know the effects on the distribution statistics of the number of repeaters in the network, the number of conversations, the amount of nonspeech data being transmitted, and other factors. Fortunately, the test bed will be designed with many monitoring points, and such information should be readily available from the TSC.

These simulation experiments will enable us to vary the relevant variables independently and to evaluate their importance. Therefore, it will also be possible for us to point to the aspects of the system that it would be most valuable to change to improve speech communication over the system. We will thus be able to provide valuable diagnostic information without requiring actual changes in the test bed itself.

We believe that the experiments described below will provide the information necessary for determining the degradation of human communication that can be expected in various digital packet networks.

Experiments 1 and 2 will be concerned with the effects of temporal distortion on high-bandwidth speech. Experiments 3, 4, and 5 will examine the combined effects of temporal distortion and bandwidth reduction of speech quality. Although speech on the actual PR net will always be bandwidth compressed, it will be important for scientific modeling and prediction purposes to distinguish the degradation effects due to temporal distortion, those due to bandwidth compression, and the interaction between the two. In our current research, we have investigated the effect of bandwidth compression alone on quality. In this proposed research program, Experiments 1 and 2 are intended to investigate the effects of temporal distortion alone on speech quality, to serve as a scientific control on the results obtained in Experiments 3, 4, and 5.

Experiment 1

Determine the effects on intelligibility and talker discrimination of interrupting high-bandwidth speech with intervals of silence.

As indicated above, a major feature of speech over a packet radio system is that, although all the speech will be delivered to the listener, the speech may be intermittently interrupted by silent intervals. Such interruptions can very seriously degrade what would otherwise be high-quality speech, that is, 200-kbit/sec PCM speech. The relative durations of uninterrupted speech and of silence will be a function of packet size and of the probability of a speech packet not being available for synthesis. Table 2 shows the duration in milliseconds of the amount of speech that would be contained in one speech packet for three different packet lengths and three different bandwidths of speech processors. On the basis of this table, we will choose durations of 30, 60, 120, and 240 msec as the length of the minimum uninterruptible speech segment in this experiment. We will then insert, with probability p , an interval of silence of duration d . Values of p will range from 0.05 to 0.5, and values of d will range from 25 to 100 msec.

This experiment will be conducted by processing MRT intelligibility tapes and ABX speaker discrimination tapes through the PDP-11/40 facility under the conditions described above. These tapes will be played to naive listeners, and the analysis of the results will enable us to plot intelligibility and talker discrimination as a function of mean duration of speech, mean duration of silence interval, and probability of silence insertion.

Table 2

MILLISECONDS OF SPEECH IN ONE PACKET

<u>Packet Size (bits)</u>	<u>Speech Processor Bandwidth (bit/sec)</u>		
	16,000	9600	2400
500	31.25	52.5	210
1000	62.5	105	420
2000	125	210	840

Experiment 2

Determine the effects on two-way interchange of information of interrupting high-bandwidth speech with intervals of silence.

The destructive effects of silence gaps on the microstructure of quality of speech as measured by single-word intelligibility and talker discrimination can potentially be largely eliminated by the introduction of a large fixed delay between talker and listener. In the extreme case, an utterance is collected at the receiver terminal in entirety, as indicated by an end-of-utterance message, packets are put into correct sequence, and the message is played to the listener. Although the availability of buffer space in the receiving terminal may place a practical limitation on the adoption of such strategies, the theoretical potential for eliminating microquality degradation exists. However, the introduction of large delays between the transmission and reception of a message could greatly degrade the efficiency of two-way exchange of information. Krauss and Bricker [1967] reported that the introduction of a 1.8-sec delay not only increases the time to complete a task but also increases the number of words needed to accomplish the task. To measure this efficiency we will compare the time required to accomplish a given two-way communication task at some fixed delay with the time required to complete it with no delay. This experiment will be conducted using the psychoacoustic consoles described in Section III. This console system enables us to measure the time required to complete an interchange task as well as the time to complete subtasks and also to precisely control the number of bits of information that must be passed from each subject to the other.

The absolute values of the parameters of this experiment cannot be specified at this time. First, we need the results of Experiment 1 to indicate what parameter values cause severe, moderate, and negligible degradation of intelligibility and talker discrimination. Second, we need measurements from the TSC to determine what fixed delays correspond to various rates and durations of interruptions. Third, we need to conduct informal tests to determine what degrees of complexity--as measured by bits transferred--adequately span an interesting set of information transfer tasks. Having obtained these results and measurements, we will do experiments that measure information interchange efficiency as a function of fixed delay for various values of task complexity, speech duration, silence duration, and probability of interruption.

Experiment 3

Determine the effects of silence interruption on speech quality when the speech has been processed by medium- and low-bandwidth reduction systems.

This experiment will use the same measurements of system merit and the same parameter values as Experiment 1. However, in this case the speech tapes will have been previously processed by medium bandwidth--CVSD, 9.6- and 16.0-kbit/sec--and low-bandwidth--LPC, 2.4-kbit--compression systems.

From our current research we know the effects of bandwidth compression alone on speech quality. From the results of Experiment 1, we will know the effects of temporal distortion alone on quality. The purpose of this experiment will be to determine the interaction between these two effects.

Experiment 4

Determine the effects of fixed delays on efficiency of information exchange when the speech has been processed by medium- and low-bandwidth compression systems.

This experiment will relate to Experiment 2 in the same way that Experiment 3 relates to Experiment 1. That is, we will be looking at information exchange after processing by bandwidth-compression systems. Again, parameter values cannot be exactly specified. However, the result of this experiment will be the specification of efficiency of two-way information exchange as a function of fixed delay, duration of uninterrupted speech, duration of silence intervals, and probability of interruption.

Experiment 5

Determine the effects on speech quality of specific conditions on the packet radio communications network.

This experiment will provide the first test of the ability of the joint TSC-SSRC effort to predict the success of human communication on a packet system of known configuration. This experiment will begin when the TSC has information on the effects of network load, geographical configuration, and so forth on parameters such as fixed delay time and probability of interruption. At that point, the SSRC will have information relating these parameters to quality as measured by intelligibility,

talker discrimination, and information exchange efficiency. From these relationships, it should be possible to predict the effects of actual packet radio parameters on speech quality. To investigate the validity of these relationships we will determine single-word intelligibility and talker discrimination scores for three to five representative configurations on the PR net. Thus the results of this experiment will determine how well we can predict speech quality over a network, given that we know the temporal-distortion and bandwidth-compression characteristics of the network.

Experiment 6

Investigate second-order effects.

This proposal has mentioned several effects that may, in fact, not occur. Examples are playing speech segments out of order and losing segments altogether. Although we believe that protocols can be found that will avoid these aberrations, this experiment makes provision for investigating such phenomena should such investigation become necessary. Parameters and parameter values will be selected to represent values that will be found in a PR net. A related problem that has been alluded to is the question of how to fill the nonspeech gaps. Possibilities are silence, white noise, the last speech frame, and a natural-sounding interruption, such as a cough. We believe that the choice of gap-filling strategy will have relatively little effect on intelligibility. However, pilot experiments will be conducted to determine whether this may be an important variable for some rates and durations of interruption.

V WORK STATEMENT

SRI will provide the facilities and the qualified personnel necessary to design and conduct a series of experiments to investigate degradations in quality of speech that may occur in the transmission of speech over packet-switched digital communication networks. These experiments are described below.

Experiment 1--Determine the effects of interruptions of high bandwidth speech on intelligibility and talker discrimination.

In this experiment, potentially very-high-quality--200-kbits/sec--digital speech will be intermittently interrupted by silent intervals. The parameters of the experiment will be minimum duration of uninterrupted speech, mean duration of silent intervals, and probability of interruption. We will measure intelligibility through MRT lists and talker discrimination through ABX tests, as a function of each of these parameters.

Experiment 2--Determine the effects of fixed delays in high-bandwidth speech conversation on two-way information exchange.

In this experiment, subjects will be asked to exchange information (precisely controlled as to number of bits of information that need to be exchanged) over a high-quality communication channel. The parameter to be varied will be the amount of delay between sending and receiving an utterance. We will measure efficiency of task accomplishment for different delays as compared to efficiency with no delay.

Experiment 3--Determine the effects of interruptions of medium- and low-bandwidth speech on intelligibility and talker discrimination.

This experiment will parallel Experiment 1, except that the speech will be processed by two medium-bandwidth (16-kbit and 9.6-kbit CVSD) and one low-bandwidth (2.4-kbit) reduction systems. The purpose of this experiment will be to determine how much degradation is due to bandwidth reduction, how much is due to temporal distortion, and how much is the result of the combination of the two effects.

Experiment 4--Determine the effects of fixed delays in medium- and low-bandwidth speech on two-way information exchange.

This experiment will parallel Experiment 2. The purpose will be to determine how much efficiency, if any, is lost in fixed delay conversations by bandwidth-compression processing. This experiment will yield measurements of two-way information exchange efficiency as a function of fixed delay time, task complexity, and type of bandwidth-reduction processing.

Experiment 5--Validate the results obtained in the previous experiments on the packet radio test bed.

Using the results of measurements made by the TSC relating network configuration and load to such parameters as interruption rate, interruption duration, and fixed delay, and using our results relating those parameters to speech intelligibility and talker discrimination, we will predict speech quality results for three to five interesting configurations of the PRnet. We will then make measurements of intelligibility and talker discrimination on the actual test bed under these configurations. These results will tell us how well we can predict the quality of speech that can be expected for a network for which temporal and bandwidth parameters are known.

Experiment 6--Investigate second-order effects.

There are several possible problems in packet radio speech communication that we believe will either be avoidable through the use of appropriate protocols or will have only slight effect on speech quality. However, the seriousness of these problems cannot be known for sure until speech communication over the packet radio system has been implemented. These problems include the possibility of speech packets being played out of order and the possibility of speech packets having to be discarded. Closely related to these problems is the possible desirability of filling gaps with other than silence. Pilot experiments will be run to determine the effects of these phenomena on speech quality.

VI MANAGEMENT PLAN

A. Supervision and Coordination

This proposed evaluation program will be conducted in the Sensory Sciences Research Center of SRI, Dr. Karl D. Kryter, Director. The principal investigator will be Richard W. Becker. Biographies of the principal personnel who will work on this project are attached. A detailed budget is included in Part Two of this proposal.

The success of this project is dependent on the actual implementation of a packet radio test bed, the implementation of speech communication on such a test bed, and the resultant measurement of statistics of transmission times, delays, and so forth under various load and geographical configurations. Thus close cooperation is required among the three projects intended to: (1) implement the packet radio test bed, (2) implement speech communication on the test bed and make measurements of relevant parameters, and (3) evaluate communication performance under various configurations. All three projects will be within the Information Sciences and Engineering Division at SRI. Informal communication on a professional level occurs frequently among the personnel of these projects. Because of the scheduling complexity involved in simultaneous development and evaluation of a novel complex system, weekly management coordination meetings will be held among the three projects.

B. Reports

Quarterly management reports, technical reports, and an annual report will be submitted. Special reports will be issued to summarize

important technical findings; when appropriate, and with the permission of the client, we will publish important results in scientific journals.

C. Schedule

The experiments described in this proposal are intended to occur in chronological order, and the design of each experiment is dependent on the results of the previous experiments. While planning of expenditure of manpower and funds must necessarily remain flexible under these circumstances, it is estimated that effort will be divided approximately equally among the first five experiments, with less time being devoted to Experiment 6. Work on these experiments can be started immediately.

VII FACILITIES

Extensive facilities for execution of this research are located at SRI. These include a DEC PDP-10, a DEC PDP-11/40 with an associated SPS-41 high-speed processor, two PDP-11/20s, and two psychoacoustic consoles controlled by a PDP-10 computer. Also available are sound-isolated rooms in close proximity to the computers to be used in the simulation experiments, AM and FM tape recorders, and facilities for presentation of tapes to up to 12 listeners per session.

Some of these computer facilities are government-furnished facilities now in use at the SRI Artificial Intelligence Center. If these facilities have sufficient capacity and if permission can be obtained, these facilities will be used for this project, and the computer budget of the proposed project will contribute to the operations costs of the installation. Otherwise, the computer budget will be used to cover communication costs and computer rental charges for similar computation services elsewhere.

The above-mentioned facilities will largely suffice for execution of this proposed research. However, capital costs of approximately \$2,000 are included for interfacing, which may be required in Experiment 4 of this research.

VIII REFERENCES

- R. Becker and F. Clarke, "Measurement of Speech Quality," SRI report, Contract DAAB 03-67-C-0070 (NSA) (Nov. 1967).
- R. Becker and K. Kryter, "Assessment of the Acceptability of Digital Speech Communication Systems," SRI report, Contract DAHC04-75-C-0008 (ARPA) (May 1975).
- R. Becker and K. Kryter, "Quality of Digital Speech Communication Systems," SRI report, Contract DAHC04-75-C-0008 (ARPA) (Jan. 1976).
- F. R. Clarke et al., "Technique for Evaluation of Speech Systems," SRI report, Contract DA 28-043 AMC-00227(E) (US Army Electronics Lab.) (April 1965).
- F. R. Clarke and R. W. Becker, "Comparison of Techniques for Discriminating among Talkers," J. Speech and Hearing Res. (Dec. 1969).
- H. H. Cohen et al., "Noise Effects, Arousal, and Human Information Processing Task Difficulty and Performance," North Carolina State University report, Contract NGL-34-002-055 (NASA) (March 1973).
- H. Dudley, "The Vocoder," Bell Lab. Record, Vol. 17, pp. 122-126 (1939).
- J. P. Egan, "Articulation Testing Methods," Laryngoscope (1948).
- G. Fairbanks, N. Guttman, and M. Miron, "Auditory Comprehension in Relation to Listening Rate and Selective Verbal Redundancy," J. Speech and Hearing Res. (1957).
- G. Fairbanks and L. Hoaglin, "An Experimental Study of the Durational Characteristics of the Voice during the Expression of Emotion," Speech Monographs (1941).
- J. Flanagan, Speech Analysis, Synthesis, and Perception (Springer Verlag, New York, 1972).
- M. Hecker and C. Williams, "Choice of Reference Conditions for Speech Preference Tests," J. Acoust. Soc. Am. (1955).

- L. Holmgren, "Speaker Recognition," Contract AFCRL-63-119 (May 1963).
- A. S. House et al., "Articulation Testing Methods: Consonantal Differentiation with a Closed Response Set," J. Acoust. Soc. Am. (1965).
- A. Huggins, "Perception of Temporally Segmented Speech," Proc. Seventh Inter Congress of Phonetic Sciences (1972).
- F. Ingeman, "Identification of the Speaker's Sex from Voiceless Fricatives," J. Acoust. Soc. (1968).
- R. Krauss and P. Bricker, "Effects of Transmission Delay and Access Delay on the Efficiency of Verbal Communication," J. Acoust. Soc. Am. (Feb. 1967).
- K. D. Kryter, K. N. Stevens, and M. Hecker, "An Evaluation of Speech Compression Systems," BBN report Contract USAF 30(602)-2235 (RADC), (March 1962).
- N. Markel, R. Eisler, and H. Reese, "Judging Personality from Dialect," J. Verbal Learning and Verbal Behavior (1967).
- G. Miller and J. Licklider, "The Intelligibility of Interrupted Speech," J. Acoust. Soc. Am. (1950).
- W. Munson and J. Karlin, "Isopreference Method for Evaluating Speech-Transmission Circuits," J. Acoust. Soc. Am. (1962).
- I. Pollack, J. Pickett, and W. Sumby, "On the Identification of Speakers by Voice," J. Acoust. Soc. Am., 26, 403 (1964).
- R. Ptacek and E. Sander, "Age Recognition from Voice," J. Speech and Hearing Res., 273 (1966).
- D. Richards and J. Swaffield, "Assessment of Speech Communication Links," Proc. Instn. Elec. Engrs., 106, Part B, No. 26, 77 (1959).
- E. Schubert and C. Parker, "Addition to Cherry's Findings on Switching Speech Between the Two Ears," J. Acoust. Soc. Am. (1955).
- W. Voiers, "Perceptual Bases of Speaker Identity," J. Acoust. Soc. Am., 36, 1065 (1964).
- W. Voiers, A. Sharpley, and C. Hehmsoth, "Research on Diagnostic Evaluation of Speech Intelligibility," Tracor report, Contract F19628-70-C-0182 (Jan. 1973).

RICHARD W. BECKER, SENIOR RESEARCH PSYCHOLOGIST
SENSORY SCIENCES RESEARCH CENTER
INFORMATION SCIENCE AND ENGINEERING DIVISION

Specialized professional competence

- Acoustic phonetics; speaker identification by human and machine; design of large-scale interactive computer systems

Representative research assignments at SRI (since 1964)

- Design and implementation of automatic natural speech understanding system
- Design and implementation of semi-automatic talker identification system
- Design and implementation of interactive speech analysis and synthesis system
- Research in characteristics used by humans in recognizing speakers
- Research in methods of evaluating quality of communication systems
- Design and implementation of real-time computer monitoring system for automatic classification of sleep stages

Other professional experience

- Research associate, Institute for Communications Research, Stanford University

Academic background

- B.S. in economics, politics, and science (1959), Massachusetts Institute of Technology; M.S. in mathematics (1962), University of Illinois; postgraduate courses in psychology and acoustic-phonetics (1959-62), University of Illinois

Publications

- Coauthor of "Comparison of Techniques for Discriminating Among Talkers," *J. Speech & Hearing Res.* (December 1969); "Natural Speech from a Computer," *Proc. 23rd ACM National Conference*; "A Model for Simulation of Attitude Change Within an Individual," *Paul J. Deutschmann Memorial Papers in Mass Communication Research* (1963)

Professional association

- Association for Computing Machinery

JAMES E. DAVIS, ENGINEERING ASSOCIATE
SENSORY SCIENCES RESEARCH CENTER
INFORMATION SCIENCE AND ENGINEERING DIVISION

Specialized professional competence

- Auditory stimulus systems, includes generating, presentation, and data recovery; field and laboratory sound recordings, both AM and FM; acoustical and special electronic laboratory requirements-- design/use and repair

Representative research assignments at SRI (since 1964)

- Speaker recognition and speech discrimination research
- Noise and speech threshold detection
- Computer analysis of aircraft flyover noise
- Psychoacoustical laboratory equipment research
- Sonic boom and aircraft flyover instrumentation for sleep and startle research

Other professional experience

- Technical associate, Sylvania Electronics Systems-West; radio frequency selectivity and digital solid-state receiver projects
- Technical associate, ITT Laboratories, Palo Alto; solid-state exchange carrier, rural subscriber and power-line voice-carrier systems

Academic background

- U.S. Naval Electronics School (1948-50); Healds School of Engineering (1952-53); Foothill College (1960-63)

PAUL D. KRYTER, DIRECTOR
SENSORY SCIENCES RESEARCH CENTER
INFORMATION SCIENCE AND ENGINEERING DIVISION

Specialized professional competence

- Psychological-physiological effects of noise on humans; physical methods for measurement and evaluation of noise; basic research on audition and speech communication

Representative research assignments at SRI (since 1965)

- Director of research program on audition and noise-induced deafness
- Director of program of laboratory and field studies on effects of environmental noise on humans
- Director of research on person identification from speech signal analysis

Other professional experience

- Head, Psychoacoustics Department, Bolt Beranek and Newman, Inc. (1957-65)
- Director, Human Resources Research Laboratories, Air Force Research and Development Command
- Assistant professor of psychology, Washington University, St. Louis, Missouri
- Research and teaching fellow, Psycho-Acoustic Laboratory and Systems Research Laboratory, Harvard University

Academic background

- B.A. in psychology (1939), Butler University; Ph.D. in psychology and physiology (1943), University of Rochester

Publications

- Author of book, *The Effects of Noise on Man*, Academic Press (1970)
- More than fifty papers in scientific and technical journals

Professional associations and honors

- Acoustical Society of America (fellow)--president; American Association for the Advancement of Science (fellow); American Psychological Association (fellow)--Council of Representatives; American Standards Association--chairman of Committee on Bioacoustics and delegate to International Organization for Standardization; British Acoustical Society; Committee on Hearing, Bioacoustics, and Biomechanics, National Academy of Sciences--Executive Council; Human Factors Society of America (fellow); Society of Engineering Psychologists--president
- Sigma Xi



STANFORD RESEARCH INSTITUTE
Menlo Park, California 94025 · U.S.A.

December 31, 1975

Proposal for Research

SRI No. ISU 75-184

EVALUATION OF SPEECH TRANSMISSION
OVER PACKET SYSTEMS

Part Two--Contractual Provisions

Prepared for:

Information Processing Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

Attention: Mr. Robert Kahn

015

APR 6 1976

APR 5 1976

Proposal for Research No. ISU 75-184

EVALUATION OF SPEECH TRANSMISSION
OVER PACKET SYSTEMS

Part Two--Contractual Provisions

I ESTIMATED TIME AND CHARGES

The estimated time required to complete this project and report its results is one year. The Institute could begin work on receipt of a fully executed contract.

Pursuant to the provisions of ASPR 16-206.2, attached is a cost estimate and support schedules in lieu of the DD Form 633-4. Also enclosed is a signed form complete as to the "Detail Description of Cost Elements."

II CONTRACT FORM

It is requested that any contract resulting from this proposal be written on a cost-plus-fixed fee basis.

III ACCEPTANCE PERIOD

This proposal will remain in effect until 31 January 1976. If consideration of the proposal requires a longer period, the Institute will be glad to consider a request for an extension of time.

December 31, 1975

COST ESTIMATE

Personnel Costs

Project Supervision, 154 man-hours at \$18.82	\$ 2,898	
Senior Professional, 1,632 man-hours at \$12.54	20,465	
Professional, 462 man-hours at \$9.59	4,431	
Support Services, 154 man-hours at \$5.41	<u>833</u>	
Total Direct Labor	\$28,627	
Payroll Burden at 30%	<u>8,588</u>	
Total Salaries and Wages		\$37,215
Overhead at 110% of Salaries and Wages		<u>40,937</u>
Total Personnel Costs		\$78,152

Direct Costs

Travel	\$ 443	
Materials and Supplies	500	
Equipment	6,800	
Computer Charges, PDP-10, 10 hours at \$288	2,880	
Communications	50	
Report Production	<u>444</u>	
Total Direct Costs		11,117
Total Estimated Cost		\$89,269
Fixed Fee		<u>7,142</u>
TOTAL ESTIMATED COST PLUS FIXED FEE		\$96,411

SCHEDULE A--DIRECT LABOR

Direct labor charges are based on the actual salaries for the staff members contemplated for the project work plus a factor of 4.4% of base salary for merit increases during the contract period of performance. Frequency of salary reviews and level of merit increases are in accordance with the Institute's Salary and Wage Payment Policy as published in Topic No. 505 of the SRI Administration Manual and as approved by the Defense Contract Administration Services Region.

SCHEDULE B--OVERHEAD AND PAYROLL BURDEN RATES

These rates have been found acceptable by the Department of Defense for billing and bidding purposes for the calendar year of 1975. We request that these rates not be specifically included in the contract, but rather that the contract provide for reimbursement at billing rates acceptable to the Contracting Officer, subject to retroactive adjustment to fixed rates negotiated on the basis of historical cost data. Included in payroll burden are such costs as vacation, holiday and sick leave pay, social security taxes, and contributions to employee benefit plans.

SCHEDULE C--MATERIALS AND SERVICES

Travel

1 trip to Washington, D.C.	\$358
Subsistence, 2 days at Washington, D.C. at \$42.50/day	<u>85</u>
Total	\$443

Materials and Supplies

\$500

Equipment

Maintenance of 2 PDP 11's	\$4,800
Maintenance of PDP 11/40, \$331/month	
Maintenance of PDP 11/10, \$69/month	
Interface for PDP 11/40	\$2,000

Computers

Computation for the proposed project will be done on a PDP-10 TENEX system such as the government-furnished system now in use in the SRI Artificial Intelligence Center. If this system has sufficient capacity and if permission can be obtained, the system will be used for this project and the computer budget of the proposed project will contribute to the operations costs of the installation. Otherwise, the computer budget will be used to cover communications costs and computer rental charges for similar computational services elsewhere.

Communications

This is an estimate of the toll charges for telephone calls during the period of performance.

SCHEDULE D - REPORT COSTS

Report costs are estimated on the basis of the number of pages of text and illustrations and the number of copies of reports required in accordance with the following rates per page.

Editing	\$ 2.55
Composition	2.50
Coordination	0.74
Proofreading	0.77
Illustration	21.96
Press and Bindery	0.022

Following is a cost breakdown of the estimated cost of report production:

Printing, 45 pages at \$6.56 per page (including editing, composition, report coordination, and proofreading)	\$295
Illustration, 5 at \$21.96 per illustration	110
Press, bindery, and photography for 1750 printed pages at \$0.022 per printed page	<u>39</u>

Total \$444

PACKET SPEECH EXPERIMENTAL PROGRAM

Proposal for Research No. ECU 78-34

Part One--Technical Proposal

31 July 1978

By: E. J. Craighill, Senior Research Engineer
Telecommunication Sciences Center

Prepared for:

Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209
Attn: Vinton G. Cerf

Approved:



Donald L. Nielson, Director
Telecommunications Sciences Center



David H. Brandin, Executive Director
Computer Science and Technology Division



CONTENTS

I SUMMARY	1
II BACKGROUND	4
A. SRI Participation In Related Areas	5
B. Packet Radio Speech Implementation	6
C. Packet Speech In Other Networks	8
III OBJECTIVES	10
IV METHOD OF APPROACH	11
A. PRnet Testing and Evaluation Using Voice Traffic	11
B. Continued Speech Software Development and Maintenance	12
C. CVSD Hardware Construction	19
V HARDWARE AND FACILITIES	21
VI WORK STATEMENT	22
VII ORGANIZATION AND SCHEDULING	23
REFERENCES	25

I SUMMARY

This proposal presents a 12-month program for continued research in digitized packet speech communication systems. SRI International participated in the initial ARPAnet speech transmission experiments and implemented a voice communication capability on the PRnet during the current year under contract N00039-76-C-0364. This experience leads us to propose research in three areas:

- (1) Continued packet radio speech evaluation and demonstration to assess the impact of wideband (CVSD) speech on the PRnet.
- (2) Implementation and testing of robust network voice conference protocols.
- (3) Integration of new voice digitization devices (codecs*), such as the (Packet Network Voice Terminal) PNVN, into the PRnet.

The motivation for this research is to determine the capability of the PRnet to support limited speech traffic, in addition to data traffic, in a tactical environment where mobile radios are used for conversations with several other individuals, possibly on other networks. This should be accomplished without a severe degradation of the primary network traffic (e.g., intercomputer logistic information). The PRnet experimentation and network capabilities study needs to be performed in a timely fashion to support the ongoing development of PRnet and internet elements, such as an end-to-end (ETE) transport control protocol (TCP) and the internet datagram protocol. Continued development of voice activity detection mechanisms and the implementation and testing of robust conference control and voice stream protocols are important tasks, because of their effect on the

*The term "codec" is used in this proposal to denote all categories of voice digitization equipment and includes both waveform tracking types (ADM) and vocoders (LPCM).

development of critical network elements and the need for more experience with and testing of realistic conferencing systems.

Using simulated and actual speech conversations, ETE testing of the PRnet can be an extremely sensitive measure of network performance and an excellent benchmark against which to compare new releases of hardware and software. In addition to sufficient throughput, digital speech transmission requires a bound on the maximum delay for each packet so as to preserve speech quality. Any delay-inducing anomaly may be discernible because speech quality degrades proportionally as the number of lost (or excessively delayed) packets increases. This situation indicates the importance of using actual speech conversations to evaluate the performance of such new network features as mobile handoff and point-to-point routing.

Voice conferences are commonly required in tactical communication. These conferences are frequently conducted under stressed circumstances where one, or possibly more, elements or nodes of the communication system may fail. It is undesirable to have such component failures result in a catastrophic loss of communication. SRI has distributed a preliminary specification of a conference control protocol that features robustness and adaptability to varying network performance, which also allows more natural interchanges between conference participants. This preliminary specification is being reviewed by the network speech compression (NSC) community and will evolve into a full implementation specification. Consequently, in cooperation with ISI, we propose to implement robust conferencing systems (including conference control and voice stream protocols) that will provide the desired controlled conferencing capability, and include reliable recovery mechanisms to ensure graceful degradation in the event of element failure.

Because of the increased number and variety of packet networks, emphasis has been placed on modularizing the protocol to achieve as much network independence as possible. The PRnet could be used to test an implementation of the protocols (using a specific network interface

module) and evaluate the performance and user effectiveness of the system in various configurations (for example, using different types of codecs, such as the PNVT). To facilitate demonstration and testing of a wide variety of conferencing environments, it would be desirable to add a Floating Point Systems programmable signal processor (AP-120) to the PRnet speech configuration. The AP-120 would provide the capability for low-rate internet linear predictive coding (LPC) experiments, for expanded demonstration facilities (including speaker verification and noise filtering), and for testing of various codec schemes.

In previous and current contracts with ARPA, we have participated in the ARPAnet speech conference demonstrations and implemented protocols for voice traffic on the PRnet. In Section II we discuss the details of that experience relevant to the proposed effort and in Section III we define the objectives of the proposed research. The proposed method of approach is presented in Section IV. In Section V we discuss hardware and facilities; in Section VI we present the work statement; and in Section VII we discuss organization and staffing.

II BACKGROUND

Military data networks need a voice capability. For certain communication applications, voice proves to be the most effective information medium when all performance factors are considered (although for speed and accuracy specifically, digital messages can be superior to voice). For example, voice communication is important in the lower echelons for real-time coordination. Packet radio networks can be designed to support the high traffic demands of this application; and they can do so with greater communication flexibility and reliability than conventional systems. As packet radio technology gets cheaper, packet speech will become economically competitive as well. Furthermore, because most real-time applications place similar demands on the communication network, packet speech systems can serve as models for other systems such as targetting, acquisition, tracking, and fire control. More specifically, packet speech offers a more practical and available vehicle for evaluating and influencing the developing packet radio technology; and communication protocols designed for packet speech could be applicable to other real-time media.

We have been actively engaged in voice communication research for some time. In particular, we have participated in the ARPA program for four years. Our focus has been on developing voice communication systems with improved voice quality and system reliability while demonstrating their feasibility and usefulness. The best method for testing complicated voice communication systems is to perform experiments under actual operating conditions so that users can have hands-on experience, and their comments and critiques can be systematically evaluated. To that end, we have implemented protocols to handle voice traffic on the ARPA-supported testbed.

A. SRI Participation In Related Areas

The advent of large computer-based communication networks (like the ARPAnet) that use advanced digital data-transmission methods has increased interest in common voice and data networks. We have been most concerned with improving the quality and reliability of integrated voice and data communication and conferencing systems using currently available hardware. We have participated in the initial demonstrations of high-quality digital speech transmission at 1 to 2 kbits/s (average) on the ARPAnet. We have also demonstrated wideband (CVSD) speech communication over the PRnet during the current contract. The goal of the latter demonstration was to assess the effect of wideband speech traffic on PRnet performance.

We have also participated in a voice satellite communication study for DCA ¹ * of several demand-assignment-multiple-access techniques for the Defense Satellite Communication System of the 1980 to 1990 era. One outcome of that study was the specification of different categories of voice communication, and the identification of the significant system design parameters that are necessary to support those categories.

The development of internet packet voice techniques must consider networks with different capabilities. The ARPAnet consists of a number of larger scale computer systems with information data bases connected by land lines. The SATnet provides communication between geographically separated places via leased satellites. The PRnet provides low-cost mobile communications to a local area via radio links.

The PRnet testbed at SRI is an excellent place to assess the capacity and quality of voice communication supported by data networks. The existing nucleus at SRI of operational hardware and software plus experienced personnel, are well-suited to performing tests and experiments, as well as for demonstrating packet voice/data systems. During the current year of our packet radio speech project, we have focused on addition of speech terminals to the Bay Area PRnet testbed.

*-----
* Numbers in brackets denote references, which can be found at the end of this proposal.

We propose to use the PRnet speech capability next year for continued testing and evaluation of the PRnet capability for supporting voice traffic, development of robust voice conference protocols, and integration of new codecs into the speech interface units.

It is, of course, a realistic goal to be able to communicate from a point in one net to a point in any other, but we shall begin by considering the PRnet. Our experience with the ARPAnet conferencing indicates that improved tactical conferencing requires system reliability, quick recovery from component failures (both hardware and software), and general robustness. We propose to implement robust voice conferencing protocols for use in the PRnet and in internet environments during the next year.

B. Packet Radio Speech Implementation

During the current contract year SRI has implemented a speech communication capability on the PRnet. Thus far, the requisite hardware and software for CVSD speech transmission have been specified and implemented, and the salient network requirements for speech traffic have been determined.

1. Available Hardware

The PRnet has three types of components: stations (the control elements), repeaters (radio store-and-forward units), and terminals. The packet radio speech implementation has concentrated on the end-user portion of the terminals called speech interface units (SIUs). The SIU is hooked to a experimental packet radio (EPR) unit and contains all user interfacing, packet forming, and ETE protocol functions required for speech conferencing. Two of the SIUs are dual-processor LSI-11s equipped with interfaces for the PRnet, CVSD voice digitization, and a user input/display. One of the processors is the high-throughput path between the CVSD and the PRnet; the other processor contains both the user interface and major control modules. Similar interfaces exist on the NSC PDP-11/40 so that it can be used as another SIU. The PDP-11/40

SIU has, in addition, more memory, two CVSD units, two 1.2 megaword disks, and a high-speed signal processor (SPS-41).

2. Completed Software

The network voice protocol (NVP) has two parts: the VOICE protocol to handle the high data-rate speech data; and the CONTROL protocol to establish and maintain connections, determine vocoder parameters, and supply status information to the user. For the PRnet, we have implemented a specific VOICE protocol compatible with the NVP specification for the ARPAnet. The software is written to run under an operating system called MOS (micro-operating system) that uses very little memory and is optimized for high throughput and memory utilization. MOS provides multiple-process scheduling (round-robin), timer control, and I/O queueing. The VOICE protocol implementation is a speech transceiver with the following basic properties (see Final Report, SRI Project 5454, 1978² for a more thorough discussion):

- (1) The speech transceiver is one reentrant process, with functionally independent (but integrated) transmitter and receiver modules, and a separate codec device driver module.
- (2) The receiver implements a real-time voice data protocol, that functions without crossnet control exchanges, for voice stream reconstruction.
- (3) The transmitter performs speech activity determination via a finite state machine algorithm, which is independent of the specific codec device in use.
- (4) The speech device driver contains all device-dependent code and interprets device-independent protocol parameters (such as interpacket interval).

Ancillary to the transceiver process, two other processes were implemented: a measurement process, and a device driver process for an alphanumeric LED (conference) information display.

As part of the current contract, the current network voice conference protocol (NVCP) was carefully reviewed; it was found that extensions can be made that will better match it to the capabilities and

characteristics of the PRnet and future internet conferencing. They include:

- . Allowing the use of broadcast packet addressing (when available).
- . Using an initial connection protocol compatible with TCP.
- . Providing for unscheduled conference formation at user initiative.
- . Permitting dynamic addition and loss of conference participants.
- . Allowing for conference partitioning into subconferences and subsequent recombining (such a condition could be forced, for example, by loss of a GATEWAY between two networks).
- . Adding more comprehensive user information displays.

3. PRnet Performance Evaluation

During the current contract, ETE evaluation of the PRnet was initiated to determine the impact of the PRnet hop transport protocol on speech communication performance. The standard tests that measure reliability (number of lost or duplicate packets), and bandwidth (number of packets per second) were expanded to determine the ETE delay characteristics. Speech transport techniques require detailed knowledge of the delay distributions (not just the average value) to preserve speech quality. This effort was especially valuable because the effects of placing high data-rate traffic on the PRnet were also observed.

C. Packet Speech In Other Networks

In addition to the NVCP protocol implemented for the ARPAnet ³, another voice conference protocol is being developed for the ARPA SATnet ⁴. Although adequate for their specific environments, these protocols rely on unique network features. The design of a conference protocol should take these differences into account, but modularize the network-dependent aspects to allow future extensions to internet conferencing.

The ARPAnet is characterized by distributed store-and-forward transmission over land lines with pipelining and some noncontention service; the distributed transmission resources can be used

simultaneously by several users if they are separated geographically. The delays are moderate and principally caused by contention for network resources, but grow for participants at remote or busy portions of the network. A natural broadcast addressing mechanism does not exist.

The SATnet is characterized by a highly efficient demand-assignment transmission, but with delays that are multiples of the satellite propagation delay (270 ms). The satellite transponder provides a natural broadcast channel so that one packet can be received by all users almost simultaneously. The satellite resources are allocated through the stream contention priority-oriented demand-allocation (C-PODA) algorithm, so a speech conference need share its stream subchannel only among its participants. A satellite network has the potential for providing a very high-capacity link, although the initial experiments have been performed with a 64-kbits/s channel.

The PRnet is characterized by relatively short delays (caused principally by retransmissions) and high capacities. The retransmissions result from either contention or failure in acquiring the rapid bit synchronization that is required for the spread-spectrum receivers. The PRnet has a natural semi-broadcast capability although the delay is not as uniform, because of multiple repeater links, as it is in the SATnet.

The PRnet offers the possibility of using carrier sensing to increase the effective channel capacity, although the presence of hidden terminals limits the potential improvement. Even so, a PRnet with hidden terminals has an advantage: users do not necessarily compete for the same resources of bandwidth and power (as they must in the SATnet) so different regions of a PRnet can be mutually non-interfering (in terms of resource contention).

III OBJECTIVES

The primary objectives of this research are to investigate the architecture and characteristics of communication protocols required to support real-time tactical applications; to assess the capability of the PRnet to support voice conferencing in a mobile tactical multinet network environment; and to determine the interface requirements for low-cost tactical speech terminals applicable to PRnet and internet environments.

IV METHOD OF APPROACH

We have been implementing a speech capability on the PRnet during the current contract. The PRnet testbed at SRI is an excellent place to test an integrated packet voice/data communication network. The existing nucleus of operational hardware and software plus experienced personnel makes us well-suited to perform tests and experiments, as well as to demonstrate the packet systems.

We propose to continue using the PRnet testbed at SRI to assess the capabilities of packet systems to support real-time tactical applications and to use a packet speech conferencing system as the test vehicle. To assess PRnet real-time performance, we propose to use the existing PRnet speech facility for continued PRnet testing and evaluation. As a part of our real-time protocol investigation, we propose to implement robust voice-conferencing protocols; this task will also enhance our PRnet performance evaluation ability by enabling more sophisticated experimental paradigms. For reasons of voice conference system compatibility and internet applicability, we also propose to integrate recently developed codecs, such as the PNVT, into the PRnet SIU.

A. PRnet Testing and Evaluation Using Voice Traffic

We propose to continue the ETE performance evaluation as new packet radio software becomes available. The usual performance measurements concentrate on reliable throughput. For speech, too much delay may be incurred for totally reliable delivery, and hence there is a more stringent requirement for packet voice communication, that is, the preservation of voice quality. The requirements of speech traffic, particularly the bound on maximum ETE delay, are good benchmarks for the performance of the PRnet. In addition, transient network events that

are not normally recorded in averaged measurements are immediately apparent during a voice conversation. We propose to use a network configuration with at least one mobile SIU in order to adequately exercise the PRnet software that we expect will be released during the next year. The anticipated software will have several features that should have an impact on real-time (voice) performance. These features will include:

- (1) Assignment of point-to-point (PTP) routes including mobile handoff. The major considerations here are assignment of low-delay routes , and minimizing of delays in traffic delivery incurred during a re-labeling.
- (2) Dual processor packet radio support. The new, faster IPR units have lower processing times and hence should lower the store-and-forward delays, thus improving throughput.

A voice link will be set up in the PRnet with at least one mobile terminal. The voice streams can be closely monitored for transient degradation and recorded, and the packet and route change history can be examined to determine the cause and extent of any degradation that might occur.

B. Continued Speech Software Development and Maintenance

Continued SIU software development is required to achieve an SIU system suitable for further PRnet testing, network conferencing demonstrations, installation in networks other than the PRnet, and eventual user environment application. Furthermore, all software installed in the SIU must be periodically maintained to preserve its compatibility with the current generation of TIU (terminal interface unit) software (most notably MOS).

Software development will be performed in two major areas: elaboration and conversion of the current SIU voice transceiver, and implementation of the voice-conferencing control protocol. These areas for software development are discussed in the following sections.

1. Voice Transceiver Development

During the current contract, a general design for network speech transceiver was conceived. This design (see Final Report, SRI Project 5454, 1978² for a more thorough discussion) was modularized to isolate all network- and codec-dependent elements from the more general elements via flexible interfaces. Except for certain network-specific elements (the network driver and protocol modules) and codec-specific elements (the codec driver module), the design is independent of the host network and local codec hardware; this permits network-independent interoperation, as well as application to most network conferencing environments. Furthermore, the design's network protocol structure allows for internet application. The following modules are included in the design:

- . The network device driver (1822 driver in the PRnet). This module handles all network input and output.
- . The ETE data protocol. This module normally performs only network addressing and raw connection management--its ETE reliability protocol functions are not used. However, because it may be desirable sometimes to use reliable conference-control streams, the reliability functions can be enabled for individual streams on request. Examples of ETE protocols are Station to PR Protocol in the PRnet environment, and the TCP and datagram protocols in an internet environment.
- . The network ETE voice protocol (to be specified in cooperation with ISI prior to the start of the new contract). This module performs voice packet stream reordering, duplicate filtering, loss detection, resynchronization, and flow control. To achieve good real-time performance, the voice protocol minimizes crossnet control transactions through its timer-based sequence windowing algorithms. The voice protocol module presents a net-independent network interface to the transmitter and receiver modules; network-specific operations are performed internally, in a manner that isolates such specificity from the other transceiver modules.
- . The voice transmitter. This module performs voice activity detection (VAD) on the encoded packet stream, does network output queueing, and accomplishes flow control (based on external information received from the voice protocol module or external control interface). Flow control can be achieved by dynamic codec rate control if appropriate codec hardware is attached; otherwise, flow control can only be accomplished by enabling or disabling codec input.

- . The voice receiver. This module performs lost packet reconstruction; receive-stream selection, interruption, and mixing (or multiplexing); canned-tone generation (e.g., ringing and busy signals); and codec-output queueing.
- . The codec driver. This module provides the routines for codec input/output, and codec parameter control (enable/disable, data-rate, etc.).
- . The external control module. This module serves as the interface between the speech transceiver and the conference control modules. It is internally interfaced to the other transceiver modules, and in response to external directives, it is responsible for directing their activity and modes of operation, as well as setting their programmable parameters.

The speech transceiver implemented under the current contract (see Section 2) is a modification of the preceding design intended to minimize memory requirements, minimize execution delays, and provide numerous internal statistics-collection facilities. This was accomplished by integrating the voice protocol, transmitter, and receiver modules into a single module; memory and execution time were thus conserved by eliminating the internal communication interfaces and integrating the separate data bases to facilitate sharing. Also note that since no real-time NVP is currently defined, a protocol-free timer-based mechanism using packet sequence space windowing was designed and implemented to fulfill essential functions of the voice protocol.

The speech transceiver was implemented in a software system environment that is now obsolete. It uses a version of the MOS that is no longer supported, and lacks certain features (since implemented) that could improve performance (such as preemptive scheduling). The transceiver also employs the predecessor of DSP, NULSPP, as its network protocol module. NULSPP is no longer supported, and its process interface is incompatible with the DSP interface.

We propose further software development that will lead to a transceiver suitable for:

- . Use in further, more sophisticated measurement paradigms (which may make use of standard TIU network modules, such as XRAY, the measurement process, and traffic generators).

- . Network voice conferencing demonstrations, under the control of a conference control module.
- . Eventual application to a user conferencing software environment.
- . Installation in diverse conferencing and multinetwork environments.

To achieve these goals, the transceiver design presented above must be realized, and its compatibility with standard TIU modules must be maintained. This will be achieved by the following sequential steps:

- (1) Useful segments of code (such as the VAD routines and CVSD driver module) will be extracted from the current transceiver to save reimplementatation, and converted to run under the current MOS and DSP.
- (2) Other transmitter and receiver module code that cannot be so extracted will be reimplemented according to the design above.
- (3) The NVP (to be specified) will be implemented in the voice protocol module. At this point, it will be possible to test the new transceiver on the PRnet, in measurement and demonstration scenarios similar to those already used by the current transceiver.
- (4) The external control module will be implemented. Using a conference control module (or keyboard input simulating this module), it will be possible to dynamically manipulate all transceiver operations and parameters at runtime, enabling actual network conferencing and more sophisticated measurement paradigms.

In reimplementing the transceiver as outlined above, consideration will be given to the future integration of codecs other than CVSD into the PRnet SIU (see Section 3). In most cases, this would entail only the addition of an appropriate set of I/O interrupt handlers. However, for future versions of the PNVT, which may contain VAD hardware, it may be necessary to circumvent some or all of the transceiver VAD processing; the VAD functions will be modularized to make this possible, and they will be conditional on the specific codec.

All the software modules discussed will be documented in order to facilitate exporting the programs to other networks or computers.

Sections of code that are processor- or network-specific will be identified.

Software maintenance will be performed as needed on all completed modules both during and after the software development process. The TIU software environment has been very dynamic to date; major changes in network software and protocols (i.e., Channel Access Protocol) have affected transceiver software. In order to preserve software compatibility and usability, a software maintenance effort will be pursued to ensure that the SIU software conforms to the most recent TIU software standards.

2. Conference CONTROL Implementation

The conference control module will implement the network conference Control Protocol; its purpose is to provide the mechanisms for local and crossnet (or internet) voice transceiver control, which serve in conference establishment, management, and termination. In cooperation with ISI and Lincoln Laboratories, the preliminary conference protocol specification prepared by SRI during the current contract ⁵ is being reviewed and updated. This protocol stresses internetting and interoperability, and use of broadcast addressing, when available; it addresses and provides mechanisms for robust operation and automatic recovery in the face of network component failure, efficient and flexible conference configuration/reconfiguration, and facile control and addition of desired conference functions.

We propose to implement a subset of the conference Control Protocol for initial test and demonstration within the PRnet. This will be useful to evaluate the general design. Performance analyses of the PRnet have revealed a throughput capacity sufficient to allow a four-party conference (two telephone extensions on the PDP-11/40 SIU) for CONTROL protocol check-out. Once the general operating features of the conference Control Protocol have been validated, additional protocol functions can be implemented as needed.

Conference control module implementation will be concurrent with voice transceiver development; the work will be scheduled to permit initial debugging and testing of the conference controller as soon as the transceiver external control module is completed.

Once a working version of the conference control module is available, SRI will conduct a live conference that will test the integrated CONTROL and VOICE protocols. Specific tests of protocol robustness will be performed by introducing artificial errors and component failures.

3. Integration of New Codecs into the SIU

The PRnet SIU software currently supports only CVSD voice digitization, and no other types of codecs have been procured for SIU application. In a move toward interoperability and eventual introduction of the SIU as a user component, it would be desirable to integrate other codecs into the SIU hardware/software configuration. We propose both of two approaches: integration of the PNVT being developed by ECI, and integration of an AP-120 (costed as an option).

The PNVT under development at ECI is intended to be a military standard digital voice dataphone. Its design calls for support of an internal codec, as well as user controls (e.g., a dial) and displays (e.g., lights). Much of the design is still fluid, and we therefore propose to cooperate with ECI to assure that the SIU is compatible with the PNVT. The most notable issues are:

- . The PNVT interfacing strategy. It is essential that the interface between the PNVT and the SIU permit the efficient transfer of digitized voice data, as well as the exchange of all necessary control signals in a manner that does not compromise the timeliness of the voice stream transfer.
- . The PNVT control protocol. The control signals exchanged between the SIU and PNVT must be specified. This specification should allow for all desirable control functions (such as codec rate, user control signals, and user display and tone generation information); it must also be general enough to permit the efficient use of the PNVT with voice communication components other than the SIU.

- . The PNVT user display. A user display that fits the space and cost requirements of the PNVT, as well as the flexibility requirements of the SIU, must be selected.
- . PNVT VAD hardware. It would be desirable for all conferencing environments, including the SIU, if some part of VAD processing take place in the PNVT. The current VAD algorithm used by the SIU consists of two independent parts: one part investigates and logically categorizes a run (packet) of voice data in terms of nonsilence content and trend; the other part is a finite state machine that makes speech activity decisions, discarding runs of silence and noise, and passing useful speech. At least the first of these functions (or a modification of it) could be performed by the PNVT; we propose to conduct further VAD strategy investigation and work with ECI to define plans and specifications for the phased incorporation of VAD hardware into the PNVT.

The PNVT as currently defined can provide up to three different voice digitizations. Each has a different impact on SIU software design. One version uses a CVSD algorithm. Since the SIU currently uses CVSD, only the device-specific mechanisms for rate control and coming-out-of-silence (COS) interrupt processing (see Final Report, SRI International Project 5454 1978) may be affected. The second digitization method uses an adaptive delta modulation (ADM) algorithm. Like CVSD, ADM-encoded speech is a synchronous bit stream; processing should be similar to that for the CVSD PNVT, except for extra changes in voice stream VAD content analysis and possibly in minor device driver operations. The third type of PNVT digitization uses LPCM. This utilizes a frame-encoded voice stream, requiring an entirely new stream-handling and buffering strategy. Additionally, voice stream VAD content analysis as currently defined is not compatible with linear predictive coding. This version of the PNVT may require incorporation of a CVSD chip, to make available a second voice stream expressly for packet content analysis. This is a powerful argument for locating content analysis in the PNVT itself--otherwise, the SIU would have to read two voice streams.

The initial models of the PNVT will use bit-serial CVSD and ADM. Circuitry for framing and VAD, to appear later, will be contained

in external modules that will plug into the bit-serial PNVTs. The new generation of SIU software will be designed (insofar as possible) to support all versions, initial and eventual, of the PNVT. Other than providing the requisite device drivers and software modifications as outlined above, future PNVT configurations with VAD hardware may require that the SIU be capable of bypassing some or all of its VAD processing, depending on the device. This requirement will be supported by the new SIU software.

We also propose an optional, second approach to integrating other codecs into the SIU: the PDP-11/40 version of the SIU can be fitted with an AP-120. The AP-120 is programmable and can support a variety of codec techniques; this would make it possible to test and demonstrate the spectrum of codec strategies without procuring and supporting a wide variety of hardware. Perhaps more importantly, the AP-120 would provide the immediate capability for internet speech experiments using low-rate LPC (in case LPCMs are not available for some time). Software for the AP-120 is currently available, and has been proven in ARPAnet speech experiments. The software runs under EPOS, an operating system related to ELF and similar to MOS. In integrating the AP-120 into the PDP-11/40 SIU, it will be necessary to convert the software to a MOS environment, but this should pose no large difficulties owing to the relationship between MOS and EPOS.

C. CVSD Hardware Construction

To facilitate the eventual integration of the PNVT into the PRnet SIU, and to improve the quality of low-level speech in the interim, we propose to make several hardware modifications to the MX-521 codecs. The MX-521s currently implement CVSD in discrete logic. This circuitry must often be adjusted to remove inherent circuit oscillations by curtailing the input dynamic range; low-level speech quality consequently suffers. We propose to replace the discrete CVSD board with a new board implementing ECI's better-performing LSI-based design for the PNVT.

Depending on the details of the final specification for the PNVT (to be determined in cooperation with ECI and ISI), it may also be desirable to add logic to the MX-521s that allows independent programmable rate control for CVSD encoding and decoding, as well as circuitry for VAD. Rate control signals would set both the CVSD sampling rate and filter bandwidths. The VAD circuitry would most likely inspect and quantify the silence/nonsilence content of the voice data stream during encoding, passing this information to the LSI-11 on request.

V HARDWARE AND FACILITIES

The current packet radio speech contract has focused on implementation of a three-node PR-net. Two LSI-11-based SIUs are operational, and is the NSC PDP-11/40, which has some additional capabilities: two CVSD extensions and a disk storage system for data or measurement logging. The PDP 11/40 is also used for down-line loading of the LSI-11 SIUs. Existing maintenance contracts for these machines should be extended to give continuous coverage.

Another hardware activity proposed for next year is the upgrading of the MX-521 CVSD units with LSI chips and VAD hardware as described in the Method of Approach, Section IV. We propose to modify all units.

VI WORK STATEMENT

SRI International will furnish the necessary personnel, equipment, and facilities, and will exert its best efforts to accomplish the following tasks:

- TASK 1: Continued Packet Radio Speech Testing and Demonstration. Each test will focus on one of the following PRnet features: PTP routing assignments including mobile handoff and IPR versus EPR performance. The data gathered during all the experiments will be analyzed.
- TASK 2: Development of the ARPAnet and PRnet Packet-Processing Software for Digitized Speech Traffic. This task will include integration of existing transceiver software with new multinet MOS, net-independent net interfaces, implementation of conference control module, continued upgrading of software to achieve the new design, and implementation of device drivers for other voice-digitization techniques (ADM and LPC).
- TASK 3: Integration of New Voice Digitization Devices into PRnet SIUs. Task 3 will entail definition of the interface for PNVT; specification and testing of the speech activity detection hardware; improvement of existing CVSD units (MX-521) for interim use; and integration and testing of low-rate devices. The work will be coordinated with ECI.
- TASK 4: Purchase and Integrate an FPS AP-120 Array Processor (costed as an option). This optional task will require installation and checkout of standard ARPAnet speech communication software and crossnet testing in cooperation with ISI.
- Task 5: Preparation of final report on Tasks 1 through 3 plus total support costs.

VII ORGANIZATION AND SCHEDULING

The proposed 12-month effort will be conducted entirely within the Information Sciences and Engineering Division of SRI. Most of the personnel who will be working on this contract are in the Telecommunications Sciences Center (TSC). Dr. D. T. Magill will serve as project supervisor and Dr. E. Craighill will be the project leader. Dr. Craighill has considerable experience with configuration design, system installation, and software development for speech communication network applications. Personnel currently working on the Packet Radio Integration Project at SRI will be used as needed for the Task 1 testing and demonstrations. Messrs. D. Rubin and A. Poggio will also assist in the implementation of the voice protocols (Task 2) and integration of new codecs (Task 3). Their expertise is in design and implementation of computer-based communication systems. We will maintain close coordination with the Packet Radio Integration Project.

Task 1 will be performed as needed when new software becomes available for testing in the PRnet.

Task 2, continued SIU software design, development, implementation, and maintenance, will commence on contract award, and proceed in accordance with the constraints imposed by availability of final specifications for the protocols involved (such as the conference control protocol) and demonstrations schedules. As tests of the implemented protocols require changes in the protocol specifications, the software will be upgraded to reflect the current protocol definitions.

Task 3 includes the design and construction of four CVSD units, which will begin at the time of contract award. Thus, the hardware can be used for the tests and experiments. Two PNVT codecs will be integrated into the SIU environment when they are delivered by ECI.

Biographies of principal SRI personnel who will work on the project are included at the end of this proposal.

REFERENCES

1. C. G. Hilborn and E. J. Craighill, et al., "Implications of Demand Assignment for Future Satellite Communication Systems," Final Report, Systems Control, Inc., Palo Alto, California (June 1977).
2. E. Craighill and D. Rubin, "Packet Speech Experimental Program Final Report," SRI International, Menlo Park, California (June 1978).
3. D. Cohen, "Specifications for the Network Voice Protocol," NSC Note No. 68 (March 1976).
4. J. Forgie, "A Proposal for SATnet Voice Conferencing," NSC Note No. 104 (January 1977).
5. D. Rubin, "Specification for a General Network Conferencing System, Voice Communication Supervisor," NSC Note No. 107 (May 1977).
6. Y. Dalal, "Broadcast Protocols in Packet Switched Computer Networks", Technical Report No. 128, Stanford University (April 1977).
7. R. Sunlin, private communication.

DONALD L. NIELSON, DIRECTOR
TELECOMMUNICATIONS SCIENCES CENTER
COMPUTER SCIENCE AND TECHNOLOGY DIVISION

Specialized professional competence

- . Application of communications technology; radio propagation; radio channel characteriziation; packet-switched digital networks; ionospheric physics; nuclear effects on communication systems

Representative research assignments at SRI (since 1959)

- . Supervision of implementation and operation of packet radio network
- . UHF radio channel measurements and characterization in urban areas
- . Study of packet radio systems
- . Ignition noise measurement and suppression studies
- . Radiation of nonsinusoidal signals
- . Application of optics to signal processing
- . Investigation of long-range transequatorial propagation at VHF
- . Development of techniques and models for predicting the effects of nuclear explosions on HF communication circuits
- . Investigation of the application of oblique-incidence sounders to HF communication
- . Study of problems in HF and VHF propagation, including ground backscatter, nuclear effects, prediction, and the use of ionospheric radars

Other professional experience

- . U.S. Army Security Agency: concerned primarily with radio propagation and communication problems

Academic background

- . B.S.E.E. (1956), University of Idaho; M.S.E.E. (1962) and Ph.D. in electrical engineering (1969), Stanford University

Publications

- . Author or coauthor of more than 20 SRI technical reports in the above research areas
- . Author or coauthor of more than 12 technical papers presented at national and international symposia and published in proceedings or technical journals

Professional associations

- . Registered Professional Engineer, California
- . Institute of Electrical and Electronics Engineers
- . U.S. URSI Commission E

EARL J. CRAIGHILL, SENIOR RESEARCH ENGINEER
TELECOMMUNICATIONS SCIENCES CENTER
COMPUTER SCIENCE AND TECHNOLOGY DIVISION

Specialized professional competence

- . Artificial intelligence; pattern recognition; applied mathematics; communication theory; computer and communication networks; time series analysis; man-machine interfaces; real-time signal processing; interactive display system design; talker identification; spoken word recognition; speech compression

Representative research assignments at SRI (since 1968)

- . Designed and implemented interactive graphic display program
- . Developed acoustical processing for automatic speech-recognition system including formant tracking, digital filters, and frequency analysis
- . Developed speech bandwidth compression algorithms
- . Designed and implemented low-bit-rate speech transmission system on ARPA Network
- . Programmed a high-speed parallel signal processor for real-time operation
- . Developed protocols for asynchronous demand-access speech communication network
- . Designed speech teleconferencing system
- . Studied the system requirements for voice transmission over a satellite communication system

Other professional experience

- . Electronics Research Laboratory, Montana State University: developed computer programs for displaying multidimensional data and analyzing classification procedures; taught applied mathematics
- . Michigan State University: developed and simulated a functional model of mammalian brain structures using nonlinear decision algorithms; applied these computerized decision algorithms to speech recognition

Academic background

- . A.B. in mathematics and physics (1963), Carroll College; M.S. in electrical engineering (1965), Montana State University; Ph.D. in electrical engineering (1971), Michigan State University

Publications

- . Coauthor of description of function model of the reticular formation, in *The Mind: Biological Approaches to Its Functions*, W. C. Corning and M. W. Balaban, eds. (1968); "Automatic Speech Recognition Based on a New Segmentation Procedure," Ph.D. thesis (August 1971); coauthor of "Human Rod Visual Delay," *Nature* (January 1968)

Honors

- . Sigma Xi

D. THOMAS MAGILL, PROGRAM MANAGER
TELECOMMUNICATIONS SCIENCES CENTER
COMPUTER SCIENCE AND TECHNOLOGY DIVISION

Specialized professional competence

- Satellite communication systems; modulation-demodulation theory; statistical estimation theory; tracking system theory; voice digitization; digital modems and equalizers

Representative research assignments at SRI (since 1970)

- Project supervisor for DCA project to develop a computer program to optimally distribute satellite power and bandwidth to frequency-division users
- Project leader for ARPA effort developing a real-time linear predictive vocoder
- Project leader for effort to find a more efficient method of encoding the speech excitation signal
- Comparison of multiple-access techniques suitable for DSCS-Phase II
- Signal design for HF band tactical radio sets
- Statistical analysis of the usage of mobile radio bands

Other professional experience

- Department manager/staff scientist, Communications Systems Department, Space and Reentry Systems Division, Philco-Ford Corporation
- Developed a unified theory for the extrinsic noise performance of tracking cross correlators
- Investigated multiple- and random-access communication systems, world-wide timing systems, and target location systems
- Conducted some of the first time-division multiple-access experiments with a hard-limiting satellite transponder
- Analyzed, designed, and directed the development of high-rate quadriphase modems
- Directed the development of two experimental ground terminals for a TDMA satellite system
- Instructed graduate courses in communication theory

Academic background

- B.S. in electrical engineering (1957), Princeton University; M.S. (1960) and Ph.D. (1964) in electrical engineering, Stanford University

Professional associations and honors

- Institute of Electrical and Electronics Engineers (Groups on Acoustics, Speech, and Signal Processing, Communications, Control Systems, and Information Theory; chairman of Information Theory Group, San Francisco Section, 1970-71; group chapter coordinator of San Francisco Section, 1970-73; chairman of Control System Society Chapter, Santa Clara Valley Section, 1975-76)
- Technical Program for WESCON (cochairman, 1973)
- B. H. Whiting Memorial Scholarship (1953-57)
- Lockheed Honors Cooperative Program (1961-63)

at the National Archives

DARRYL E. RUBIN, RESEARCH ENGINEER
TELECOMMUNICATIONS SCIENCES CENTER
COMPUTER SCIENCE AND TECHNOLOGY DIVISION

Specialized professional competence

- . Computer networking and communication protocols; computer operating system design and systems programming; digital logic and micro-computer hardware design

Professional experience

- . Designed and implemented a generalized interhost file transfer package for use on a packet-switched computer network (ARPANET)
- . Designed and implemented software for a stand-alone portable terminal system for direct connection to packet-switched computer networks (ARPANET and PRnet)
- . Designed and implemented software for real-time packetized speech conferencing over a computer network (PRnet)

Academic background

- . B.S. Biology, Stanford University
- . Currently working on M.S.C.E., Stanford University

Academic honors

- . Phi Beta Kappa

Publications

- . "Telnet under Single-Connection TCP Specification," Stanford University Digital Systems Laboratory, Technical Note #76 (February 1976)

Professional associations

- . Institute of Electrical and Electronics Engineers

ANDREW A. POGGIO, SYSTEMS PROGRAMMER
TELECOMMUNICATIONS SCIENCES CENTER
COMPUTER SCIENCE AND TECHNOLOGY DIVISION

Specialized professional competence

- . Design and implementation of interactive computer software systems, computer languages, and computer network communication protocols

Representative research assignments at SRI (since 1975)

- . Designed and implemented the latest version of the oN-Line System (NLS), an interactive system for creating, viewing, and processing information maintained in a hierarchically structured form
- . Designed and implemented a system for high-quality human/software interaction (called the Front End) and a high-level language (called the Command Meta Language, CML) to describe such interaction
- . Designed and implemented software modules supporting several computer-to-computer communications protocols

Other professional experience

- . Electrical Engineering and Computer Science Department, University of California (Berkeley): designed and implemented an interactive system supporting a real-time simulation of medical experimentation in a specific context
- . Computer Science Department, University of California (Berkeley): implemented a generalized compiler-compiler system
- . Lawrence Hall of Science (Berkeley): designed and implemented an interactive system interfacing user commands to specialized hardware

Academic background

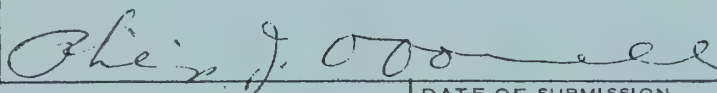
- . B.S. in computer science (1973) and M.S. in electrical engineering and computer science (1975), University of California (Berkeley)

Publications

- . Coauthor of "User Interface System for a Computer Network Marketplace," (December 8, 1976) SRI-ARC Catalog Item 28913; "Stand Alone Front End Programmer's Guide," (August 3, 1976) SRI-ARC Catalog Item 28451
- . Coauthor of several articles on the design and implementation of a system for simulation of medical experimentation and on the results obtained therefrom (1975)

Honors

- . University of California Alumni Scholar

a. PURCHASED PARTS					
b. SUBCONTRACTED ITEMS					
c. OTHER - (1) RAW MATERIAL					
(2) YOUR STANDARD COMMERCIAL ITEMS					
(3) INTERDIVISIONAL TRANSFERS (At other than cost)					
TOTAL DIRECT MATERIAL					
2. MATERIAL OVERHEAD 3 (Rate % X \$ base =)					
3. DIRECT LABOR (Specify)	ESTIMATED HOURS	RATE/HOUR	EST COST (\$)		
TOTAL DIRECT LABOR					
4. LABOR OVERHEAD (Specify department or cost center) 3	O.H. RATE	X BASE =	EST COST (\$)		
TOTAL LABOR OVERHEAD					
5. SPECIAL TESTING (Including field work at Government installations)			EST COST (\$)		
TOTAL SPECIAL TESTING					
6. SPECIAL EQUIPMENT (If direct charge) (Itemize on Exhibit A)					
7. TRAVEL (If direct charge) (Give details on attached Schedule)			EST COST (\$)		
a. TRANSPORTATION					
b. PER DIEM OR SUBSISTENCE					
TOTAL TRAVEL					
8. CONSULTANTS (Identity - purpose - rate)			EST COST (\$)		
TOTAL CONSULTANTS					
9. OTHER DIRECT COSTS (Itemize on Exhibit A)					
10. TOTAL DIRECT COST AND OVERHEAD					
11. GENERAL AND ADMINISTRATIVE EXPENSE (Rate % of cost element Nos.) 3					
12. ROYALTIES 4					
13. TOTAL ESTIMATED COST			See Part Two--Contractual		
14. FEE OR PROFIT			Provisions No. ISU 77-117		
15. TOTAL ESTIMATED COST AND FEE OR PROFIT					
This proposal is submitted for use in connection with and in response to (Describe RFP, etc.)					
- - -					
and reflects our best estimates as of this date, in accordance with the instructions to offerors and the footnotes which follow.					
TYPED NAME AND TITLE Philip J. O'Donnell Manager Contract Administration			SIGNATURE 		
NAME OF FIRM SRI INTERNATIONAL			DATE OF SUBMISSION 1977 AUG 22		

DD FORM 633-4
1 APR 68

REPLACES EDITION OF 1 JAN 67, WHICH IS OBSOLETE.

EXHIBIT A - SUPPORTING SCHEDULE (*Specify. If more space is needed, use blank sheets*)

[illegible]

1. HAVE THE DEPARTMENT OF DEFENSE, NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, OR THE ATOMIC ENERGY COMMISSION PERFORMED ANY REVIEW OF YOUR ACCOUNTS OR RECORDS IN CONNECTION WITH ANY OTHER GOVERNMENT PRIME CONTRACT OR SUBCONTRACT WITHIN THE PAST TWELVE MONTHS?

☒ YES ☐ NO *If yes, identify below.*

NAME AND ADDRESS OF REVIEWING OFFICE (Include ZIP Code)

Defense Contract Audit Agency, SRI Resident

TELEPHONE NUMBER/EXTENSION

415 326-6200, X2089

II. WILL YOU REQUIRE THE USE OF ANY GOVERNMENT PROPERTY IN THE PERFORMANCE OF THIS PROPOSED CONTRACT?

☒ YES ☐ NO *If yes, identify on a separate page.* See Part Two--Contractual Provisions No.

III. DO YOU REQUIRE GOVERNMENT CONTRACT FINANCING TO PERFORM THIS PROPOSED CONTRACT? ISU 77-117

☐ YES ☒ NO if yes, identify: ☐ ADVANCE PAYMENTS ☐ PROGRESS PAYMENTS OR ☐ GUARANTEED LOANS

IV. DO YOU NOW HOLD ANY CONTRACT (or, do you have any independently financed (IR & D) projects) FOR THE SAME OR SIMILAR WORK CALLED FOR BY THIS PROPOSED CONTRACT? ☐ YES ☒ NO

☒ YES ☐ NO If yes, identify Contract No. N00039-78-C-0384,
Naval Electronics Systems Command

V. DOES THIS COST SUMMARY CONFORM WITH THE COST PRINCIPLES SET FORTH IN ASPR, SECTION XV (See 3-807.2 (c) (2))?

☒ YES ☐ NO If no, explain on a separate page.

INSTRUCTIONS TO OFFERORS

1. The purpose of this form is to provide a standard format by which the offeror submits to the Government a summary of incurred and estimated cost (*and attached supporting information*) suitable for detailed review and analysis. Prior to the award of a contract resulting from this proposal the offeror shall, under the conditions stated in ASPR 3-807.3, be required to submit a Certificate of Current Cost or Pricing Data (see ASPR 3-807.3(e) and 3-807.4).

2. As part of the specific information required by this form, the offeror must submit with this form, and clearly identify as such, cost or pricing data (that is, data which is verifiable and factual and otherwise as defined in ASPR 3-807.3(e)). In addition, he must submit with this form any information reasonably required to explain the offeror's estimating process, including:

a. the judgmental factors applied and the mathematical or other methods used in the estimate including those used in projecting from known data, and

b. the contingencies used by offeror in his proposed price.

3. When attachment of supporting cost or pricing data to this form is impracticable, the data will be specifically identified and described (*with schedules as appropriate*), and made available to the contracting officer or his representative upon request.

4. The format for the "Cost Elements" is not intended as rigid requirements. These may be presented in different format with the prior approval of the contracting officer if required for more effective and efficient presentation. In all other respects this form will be completed and submitted without change.

5. By submission of this proposal, offeror, if selected for negotiation, grants to the contracting officer, or his authorized representative, the right to examine, for the purpose of verifying the cost or pricing data submitted, those books, records, documents and other supporting data which will permit adequate evaluation of such cost or pricing data, along with the computations and projections used therein. This right may be exercised in connection with any negotiations prior to contract award.

FOOTNOTES

1 Enter in this column those necessary and reasonable costs which in the judgment of the offeror will properly be incurred in the efficient performance of the contract. When any of the costs in this column have already been incurred (e.g., on a letter contract or change order), describe them on an attached supporting schedule. Identify all sales and transfers between your plants, divisions, or organizations under a common control, which are included at other than the lower of cost to the original transferor or current market price.

3 Indicate the rates used and provide an appropriate explanation. Where agreement has been reached with Government representatives on the use of forward pricing rates, describe the nature of the agreement. Provide the method of computation and application of your overhead expense, including cost breakdown and showing trends and budgetary data as necessary to provide a basis for evaluation of the reasonableness of proposed rates.

I. HAVE THE DEPARTMENT OF DEFENSE, NATIONAL AERONAUTICS AND SPACE ADMINISTRATION, OR THE ATOMIC ENERGY COMMISSION PERFORMED ANY REVIEW OF YOUR ACCOUNTS OR RECORDS IN CONNECTION WITH ANY OTHER GOVERNMENT PRIME CONTRACT OR SUBCONTRACT WITHIN THE PAST TWELVE MONTHS?

☒ YES ☐ NO If yes, identify below.

NAME AND ADDRESS OF REVIEWING OFFICE (Include ZIP Code)

Defense Contract Audit Agency, SRI Resident

TELEPHONE NUMBER/EXTENSION

415 326-6200, X2089

II. WILL YOU REQUIRE THE USE OF ANY GOVERNMENT PROPERTY IN THE PERFORMANCE OF THIS PROPOSED CONTRACT?

☒ YES ☐ NO If yes, identify on a separate page. See Part Two--Contractual Provisions No.

III. DO YOU REQUIRE GOVERNMENT CONTRACT FINANCING TO PERFORM THIS PROPOSED CONTRACT? ISU 77-117

☐ YES ☒ NO if yes, identify: ☐ ADVANCE PAYMENTS ☐ PROGRESS PAYMENTS OR ☐ GUARANTEED LOANS

IV. DO YOU NOW HOLD ANY CONTRACT (or, do you have any independently financed (IR & D) projects) FOR THE SAME OR SIMILAR WORK CALLED FOR BY THIS PROPOSED CONTRACT?

☒ YES ☐ NO If yes, identify Contract No. N00039-76-C-0364, Naval Electronics Systems Command

V. DOES THIS COST SUMMARY CONFORM WITH THE COST PRINCIPLES SET FORTH IN ASPR, SECTION XV (See 3-807.2 (c) (2))?

☒ YES ☐ NO If no, explain on a separate page.

INSTRUCTIONS TO OFFERORS

1. The purpose of this form is to provide a standard format by which the offeror submits to the Government a summary of incurred and estimated cost (and attached supporting information) suitable for detailed review and analysis. Prior to the award of a contract resulting from this proposal the offeror shall, under the conditions stated in ASPR 3-807.3, be required to submit a Certificate of Current Cost or Pricing Data (see ASPR 3-807.3(e) and 3-807.4).

2. As part of the specific information required by this form, the offeror must submit with this form, and clearly identify as such, cost or pricing data (that is, data which is verifiable and factual and otherwise as defined in ASPR 3-807.3(e)). In addition, he must submit with this form any information reasonably required to explain the offeror's estimating process, including:

- a. the judgmental factors applied and the mathematical or other methods used in the estimate including those used in projecting from known data, and
- b. the contingencies used by offeror in his proposed price.

3. When attachment of supporting cost or pricing data to this form is impracticable, the data will be specifically identified and described (with schedules as appropriate), and made available to the contracting officer or his representative upon request.

4. The format for the "Cost Elements" is not intended as rigid requirements. These may be presented in different format with the prior approval of the contracting officer if required for more effective and efficient presentation. In all other respects this form will be completed and submitted without change.

5. By submission of this proposal, offeror, if selected for negotiation, grants to the contracting officer, or his authorized representative, the right to examine, for the purpose of verifying the cost or pricing data submitted, those books, records, documents and other supporting data which will permit adequate evaluation of such cost or pricing data, along with the computations and projections used therein. This right may be exercised in connection with any negotiations prior to contract award.

FOOTNOTES

1 Enter in this column those necessary and reasonable costs which in the judgment of the offeror will properly be incurred in the efficient performance of the contract. When any of the costs in this column have already been incurred (e.g., on a letter contract or change order), describe them on an attached supporting schedule. Identify all sales and transfers between your plants, divisions, or organizations under a common control, which are included at other than the lower of cost to the original transferee or current market price.

2 When space in addition to that available in Exhibit A is required, attach separate pages as necessary and identify in this "Reference" column the attachment in which information supporting the specific cost element may be found. No standard format is prescribed; however, the cost or pricing data must be accurate, complete and current, and the judgment factors used in projecting from the data to the estimates must be stated in sufficient detail to enable the contracting officer to evaluate the proposal. For example, provide the basis used for pricing materials such as by vendor quotations, shop estimates, or invoice prices; the reason for use of overhead rates which depart significantly from experienced rates (reduced volume, a planned major rearrangement, etc.); or justification for an increase in labor rates (anticipated wage and salary increases, etc.). Identify and explain any contingencies which are included in the proposed price, such as anticipated costs of rejects and defective work, or anticipated technical difficulties.

3 Indicate the rates used and provide an appropriate explanation. Where agreement has been reached with Government representatives on the use of forward pricing rates, describe the nature of the agreement. Provide the method of computation and application of your overhead expense, including cost breakdown and showing trends and budgetary data as necessary to provide a basis for evaluation of the reasonableness of proposed rates.

4 If the total royalty cost entered here is in excess of \$250 provide on a separate page (or on DD Form 783, Royalty Report) the following information on each separate item of royalty or license fee: name and address of licensor; date of license agreement; patent numbers, patent application serial numbers, or other basis on which the royalty is payable; brief description, including any part or model numbers of each contract item or component on which the royalty is payable; percentage or dollar rate of royalty per unit; unit price of contract item; number of units; and total dollar amount of royalties. In addition, if specifically requested by the contracting officer, a copy of the current license agreement and identification of applicable claims of specific patents shall be provided.

5 Provide a list of principal items within each category indicating known or anticipated source, quantity, unit price, competition obtained, and basis of establishing source and reasonableness of cost.

